Multivariate Generalised Linear Mixed Models via sabreStata (Sabre in Stata)

Rob Crouchley
r.crouchley@lancaster.ac.uk
Centre for e-Science
Lancaster University

Dave Stott
d.stott@lancaster.ac.uk
Centre for e-Science
Lancaster University

John Pritchard
j.pritchard@lancaster.ac.uk
Centre for e-Science
Lancaster University

version 1

Contents

Preface

The main aims of this book are: to provide an introduction to the principles of modelling as applied to longitudinal data from panel and related studies with the necessary statistical theory; and to describe the application of these principles to the analysis of a wide range of examples using the Sabre software (`http://sabre.lancs.ac.uk/`).

This material on multivariate generalised linear mixed models arises from the activities at the Economic and Social Research Council (ESRC) funded Colaboratory for Quantitative e-Social Science (CQeSS) at Lancaster University over the period 2003-2008. Sabre is a program for the statistical analysis of multi-process event/response sequences. These responses can take the form of binary, ordinal, count and linear recurrent events. The response sequences can also be of different types (e.g. linear (wages) and binary (trade union membership)). Such multi-process data are common in many research areas, e.g. in the analysis of work and life histories from the British Household Panel Survey or the German Socio-Economic Panel Study where researchers often want to disentangle state dependence (the effect of previous responses or related outcomes) from any omitted effects that might be present in recurrent behaviour (e.g. unemployment). Understanding of the need to disentangle these generic substantive issues dates back to the study of accident proneness in the 1950s and has since been discussed in many applied areas, including consumer behaviour and voting behaviour.

Sabre can also be used to model collections of single sequences such as may occur in medical trials on the number of headaches experienced over a sequence of weeks, or in single-equation descriptions of cross-sectional clustered data such as the educational attainment of children in schools.

Sabre is available in three forms: (1) stand-alone, (2) the R plugin, (3) the Stata plugin (as discussed here) for Windows and Linux PCs. The stand-alone version and the R plugin versions can be deployed in parallel on high performance computers (HPCs) or computational grids running Linux.

The class of models that can be estimated by Sabre may be termed Multivariate Generalised Linear Mixed Models (MGLMMs). These models have special features to help them disentangle state dependence from the incidental parameters (omitted or unobserved effects). The incidental parameters can be treated as random or fixed. The random effects models can be estimated with standard Gaussian quadrature or adaptive Gaussian quadrature. Even though the linear model integral has a closed form solution, we do not use it. Current computational facilities on many desktop computers often make the delay involved in using numerical integration for the linear model negligible for many small to medium-sized data sets. For large problems, we can always use parallel Sabre on a HPC or computational grid. 'End effects' can also be added to the models to accommodate 'stayers' or 'non-susceptibles'. The fixed effects algorithm we have developed uses code for large sparse matrices from the Harwell Subroutine Library, see `http://www.cse.scitech.ac.uk/nag/hsl/`.

Also included in Sabre is the option to undertake all the calculations using increased accuracy. Numerical underflow and overflow often occur in the estimation process for models with incidental parameters. We suppose that many of the alternative software systems truncate their calculations without informing the user when this happens as there is little discussion of this in their respective user manuals.

This book is written in a way that we have found appropriate for some of our short courses. The book starts with the simple linear two level random effects model and gradually adds complexity with the two level random effects binary and Poisson response models. We then review the generalised linear model notation before illustrating a range of more substantively appropriate random effects models, e.g. the three-level model, multivariate, endpoint, event history and state dependence models. The MGLMMs are estimated using either standard Gaussian quadrature or adaptive Gaussian quadrature. The book also compares two level fixed and random effects linear models. Additional information on Sabre commands, quadrature, model estimation and endogenous variables are included in Appendix A.

Appendix B reviews some of the Stata commands needed to set up the data for analysis using the Sabre in Stata plugin.

A separate booklet entitled "Exercises for sabreStata (Sabre in Stata)" is available from `http://sabre.lancs.ac.uk/`. This booklet contains the small data sets and exercises that have been written to accompany this book. These exercises will run quickly on a desktop PC. To distinguish the different types of exercise, we use a variety of suffixes. The 'C' suffix stands for Cross-sectional, the 'L' suffix stands for Longitudinal, the '3L' suffix stands for Three Level models; the 'FO' suffix stands for First Order in state dependence models, the 'EP' suffix stands for models which include Endpoints and the 'FE' suffix stands for Fixed Effects. Some medium sized and large data sets for testing deployment of Sabre on a Grid are available from `http://sabre.lancs.ac.uk/`.

Drafts of the chapters of this book were developed and revised in the process of preparing and delivering short courses in 'Statistical Modelling using Sabre', 'Multilevel Modelling' and 'Event History Analysis' given at CQeSS and the Department of Mathematics and Statistics at Lancaster University and elsewhere. We are grateful to many of the students on these courses from a range of backgrounds (e.g. computational science, social science) whose comments and criticisms improved these early drafts. We think that the book should serve as a self-teaching manual for the applied quantitative social scientist.

If you have any suggestions as to how this book could be improved, for instance by the addition of other material, could you please let us know via the Sabre mailing list, `sabre@lancaster.ac.uk`.

Acknowledgements

Many thanks to Steve Bradley, Colin Green (Economics, Lancaster) and Martyn Andrews (Economics, Manchester) for providing the motivation for the Stata version of Sabre. Thanks to Iraj Kazemi for helping to draft the material in the first 5 Chapters of this book. Thanks to Richard Davies for inspiring the early development of Sabre (Poisson and logit models with endpoints) and for the work on the TRAMSS site,
http://tramss.data-archive.ac.uk/documentation/migration/migpag0.htm#Top.

Many thanks to Dan Grose for writing the Stata `do` file that helped simplify the sabreStata commands and for the initial work on the Stata menu system for Sabre (not yet publicly available). Dave Stott and John Pritchard did all of the recent development work on Sabre. Dave wrote the standard Gaussian and adaptive Gaussian quadrature algorithms. John wrote the algorithm for manipulating the large sparse matrices used by the fixed effect estimator. John also wrote the procedures that enable Sabre to go parallel on multiple processors.

This work was supported by the ESRC research grants RES-149-28-1003: The Colaboratory for Quantitative e-Social Science (E-Social Science Centre Lancaster Node) and RES-149-25-0010: An OGSA Component-based Approach to Middleware for Statistical Modelling. Rob Crouchley was the principal applicant on both grants.

Chapter 1

Linear Models I

1.1 Random Effects ANOVA

The simplest multilevel model is equivalent to a one-way analysis of variance with random effects in which there are no explanatory variables. This model contains only random variation between the level-2 units and random variation within level-2 units. This model is useful as a conceptual building block in multilevel modelling as it possesses only the explicit partition of the variability in the data between the two levels.

Suppose that y_{ij} denotes the response variable for level-1 unit i within level-2 unit j, then the simplest multilevel model can be expressed as a model where the response variable is the sum of a random intercept for the level-2 units j, β_{0j}, and the residual effect for the level-1 units i within these level-2 units, ε_{ij}:

$$y_{ij} = \beta_{0j} + \varepsilon_{ij}.$$

Assuming the ε_{ij} have zero means, the intercept β_{0j} can be thought of as the mean of level-2 unit or group j. Groups with a high value of β_{0j} tend to have, on average, high responses whereas groups with a low value of β_{0j} tend to have, on average, low responses. The level-2 equation also has no predictors in its simplest form:

$$\beta_{0j} = \gamma_{00} + u_{0j},$$

where β_{0j} is the dependent variable, γ_{00} is the level-2 intercept, and u_{0j} is the level-2 error with mean 0. In this equation, γ_{00} represents the grand mean or the mean of the group-specific intercepts and u_{0j} represents the deviation of each group-specific mean from the grand mean. When the average deviation is large, there are large group differences.

Rewriting the two equations as a single equation, we have

$$y_{ij} = \gamma_{00} + u_{0j} + \varepsilon_{ij}$$

where γ_{00} is the population grand mean, u_{0j} is the specific effect of level-2 unit j, and ε_{ij} is the residual effect for level-1 unit i within this level-2 unit. In other words, level-2 unit j has the 'true mean' $\gamma_{00} + u_{0j}$, and each measurement of a level-1 unit within this level-2 unit deviates from this true mean by some value, called ε_{ij}. Level-2 units differ randomly from one another, which is reflected by the fact that u_{0j} is a random variable and that this type of model is called a 'random effects model'. Some level-2 units have a high (low) true mean, corresponding to a high (low) value of u_{0j} while other level-2 units have a true mean close to the average, corresponding to a value of u_{0j} close to zero. .

It is assumed that the random variables u_{0j} and ε_{ij} are mutually independent, the group effects u_{0j} having population mean 0 and variance $\sigma^2_{u_0}$ (the population between-group variance), and the residuals ε_{ij} having mean 0 and variance σ^2_ε (the population within-group variance). For example, if level-1 units are children and level-2 units are schools, then the within-group variance is the variance between children within the schools about the true school mean, while the between-group variance is the variance between the schools' true means.

The one-way analysis of variance examines the deviations of group means from the grand mean. Here, it is assumed that the group means, represented by $\mu_{ij} = \gamma_{00} + u_{0j}$ and, thus, their deviations are varying randomly. Therefore, this model is equivalent to the random effects ANOVA model, for further details see e.g. Hsiao (1986), Rabe-Hesketh and Skrondal (2005) and Wooldridge (2006).

1.2 The Intraclass Correlation Coefficient

A basic measure for the degree of dependency in grouped observations is the intraclass correlation coefficient. The term 'class' is conventionally used here and refers to the level-2 units in the classification system under consideration. There are, however, several definitions of this coefficient, depending on the assumptions about the sampling design.

Consider the model $y_{ij} = \gamma_{00} + u_{0j} + \varepsilon_{ij}$. The total variance of y_{ij} can be decomposed as the sum of the level-2 and level-1 variances,

$$var(y_{ij}) = var(u_{0j}) + var(\varepsilon_{ij}) = \sigma^2_{u_0} + \sigma^2_\varepsilon.$$

The covariance between responses of two level-1 units (i and i', with $i \neq i'$) in the same level-2 unit j is equal to the variance of the contribution u_{0j} that is shared by these level-2 units,

$$cov(y_{ij}, y_{i'j}) = var(u_{0j}) = \sigma^2_{u_0}.$$

The correlation between values of two randomly drawn level-1 units in the same, randomly drawn, level-2 unit is given by

$$\rho(y_{ij}, y_{i'j}) = \frac{\sigma^2_{u_0}}{\sigma^2_{u_0} + \sigma^2_\varepsilon}.$$

This parameter is called *the intraclass correlation coefficient* or the intra-level-2-unit correlation coefficient. It is seen that the coefficient ρ is:

$$\rho = \frac{\text{population variance between level-2 units}}{\text{total variance}}.$$

The intraclass correlation coefficient ρ measures the proportion of the variance in the outcome that is between the level-2 units. We note that the true correlation coefficient ρ is restricted to take non-negative values, i.e. $\rho \geq 0$. The existence of a positive intraclass correlation coefficient, i.e. $\rho > 0$, resulting from the presence of more than one residual term in the model, means that traditional estimation procedures such as Ordinary Least Squares (that is, assuming $\sigma^2_{u_0} = 0$), which are used in multiple regression with fixed effects, are inapplicable.

- Note that, conditional on being in group j

$$E(\overline{y}_{.j}|\beta_{0j}) = \beta_{0j},$$
$$Var(\overline{y}_{.j}|\beta_{0j}) = \frac{\sigma^2_\varepsilon}{n_j}.$$

- But across the population

$$E(\overline{y}_{.j}) = \gamma_{00},$$
$$Var(\overline{y}_{.j}) = \sigma^2_{u_0} + \frac{\sigma^2_\varepsilon}{n_j}.$$

Features of note:

1. The unconditional mean is equal to the expectation of the mean conditional mean.
2. The unconditional variance is equal to the mean of the conditional variance plus the variance of the conditional mean.

1.3 Parameter Estimation by Maximum Likelihood

There are three kinds of parameters that can be estimated:

1. The regression parameters: in this case there is only one, the constant: γ_{00}
2. The variance components: $\sigma^2_{u_0}$ and σ^2_e.
3. Random effects: β_{0j} or, equivalently, combined with γ_{00}: u_{0j}.

The model is

$$y_{ij} = \mu_{ij} + \varepsilon_{ij},$$
$$\mu_{ij} = \gamma_{00} + u_{0j}.$$

The likelihood function is given by

$$L\left(\gamma_{00}, \sigma_{\varepsilon}^2, \sigma_{u_0}^2 | \mathbf{y}\right) = \prod_j \int_{-\infty}^{+\infty} \prod_i g\left(y_{ij}|u_{0j}\right) f\left(u_{0j}\right) du_{0j},$$

where

$$g\left(y_{ij}|u_{0j}\right) = \frac{1}{\sqrt{2\pi}\sigma_{\varepsilon}} \exp\left(-\frac{\left[y_{ij} - \mu_{ij}\right]^2}{2\sigma_{\varepsilon}^2}\right),$$

and

$$f\left(u_{0j}\right) = \frac{1}{\sqrt{2\pi}\sigma_{u_0}} \exp\left(-\frac{u_{0j}^2}{2\sigma_{u_0}^2}\right).$$

Maximization of the likelihood function over the parameter space gives MLEs for $\theta = \left(\gamma_{00}, \sigma_{\varepsilon}^2, \sigma_{u_0}^2\right)$. Sabre evaluates the integral $L\left(\gamma_{00}, \sigma_{\varepsilon}^2, \sigma_{u_0}^2 | \mathbf{y}\right)$ for the linear model using normal Gaussian quadrature or adaptive Gaussian quadrature (numerical integration). Note that the random effects u_{0j} are latent variables rather than statistical parameters, and therefore are not estimated as an integral part of the statistical parameter estimation. Nevertheless, they may be predicted by a method known as *empirical Bayes estimation* which produces so-called *posterior means*. The basic idea of this method is that u_{0j} can be predicted (or estimated) by combining two kinds of information:

1. the data from group j,
2. the fact that the unobserved u_{0j} is a random variable with mean 0 and variance $\sigma_{u_0}^2$.

In other words, data information is combined with population information.

The posterior means for the level-2 residual u_{0j} are given by

$$\widehat{u}_{0j} = E\left(u_{0j}|\mathbf{y}, \theta\right) = \frac{\sigma_{u_0}^2}{\sigma_{u_0}^2 + \sigma_{\varepsilon}^2/n_j}\left(\overline{y}_{.j} - \overline{y}\right),$$

where θ are the model parameters, see Goldstein (1987).

The estimate for the intercept β_{0j} will be the same as the estimate for u_{0j} plus γ_{00}. Note that, if we used only group j, β_{0j} would be estimated by the group mean,

$$\hat{\beta}_{0j} = \overline{y}_{.j}.$$

If we looked only at the population, we would estimate β_{0j} by its population mean, γ_{00}. This parameter is estimated by the overall mean,

$$\hat{\gamma}_{00} = \overline{y}.$$

If we combine the information from group j with the population information, the combined estimate for β_{0j} is a weighted average of the two previous estimates:

$$\hat{\beta}_{0j}^{EB} = w_j \hat{\beta}_{0j} + (1 - w_j)\, \hat{\gamma}_{00},$$

where $w_j = \dfrac{\sigma_{u_0}^2}{\sigma_{u_0}^2 + \sigma_{\varepsilon}^2 / n_j}$. The factor w_j is often referred to as a 'shrinkage factor' since it is always less than or equal to one. As n_j increases this factor tends to one, and as the number of level-1 units in a level-2 unit decreases the factor becomes closer to zero. In practice we do not know the true values of the variancess $\sigma_{u_0}^2$ and σ_{ε}^2, and we substitute estimated values to obtain $\hat{\beta}_{0j}^{EB}$.

1.4 Regression with level-2 effects

In multilevel analysis the level-2 unit means (group means for explanatory variables) can be considered as an explanatory variable. A level-2 unit mean for a given level-1 explanatory variable is defined as the mean over all level-1 units, within the given level-2 unit. The level-2 unit mean of a level-1 explanatory variable allows us to express the difference between within-group and between-group regressions. The within-group regression coefficient expresses the effect of the explanatory variable within a given group; the between-group regression coefficient expresses the effect of the group mean of the explanatory variable on the group mean of the response variable. In other words, the between-group regression coefficient is just the coefficient in a regression analysis for data that are aggregated (by averaging) to the group level.

A cross-sectional example will be demonstrated.

1.5 Example C1. Linear Model of Pupil's Maths Achievement

The data we use in this example are a sub-sample from the 1982 High School and Beyond Survey (Raudenbush and Bryk, 2002), and include information on 7,185 students nested within 160 schools: 90 public and 70 Catholic. Sample sizes vary from 14 to 67 students per school.

1.5.1 Reference

Raudenbush, S.W., Bryk, A.S., 2002, Hierarchical Linear Models, Thousand Oaks, CA. Sage

1.5.2 Data description for `hsb.dta`

Number of observations (rows): 7185
Number of level-2 cases: 160

1.5.3 Variables

`school`: school identifier
`student`: student identifier
`minority`: 1 if student is from an ethnic minority, 0 = other
`gender`: 1 if student is female, 0 otherwise
`ses`: a standardized scale constructed from variables measuring parental education, occupation, income and, socio-economic status
`meanses`: mean of the SES values for the students in this school
`mathach`: a measure of the students' mathematics achievement
`size`: school enrolment
`sector`: 1 if school is from the Catholic sector, 0 = public
`pracad`: proportion of students in the academic track
`disclim`: a scale measuring disciplinary climate
`himnty`: 1 if more than 40% minority enrolment, 0 if less than 40%

school	student	minority	gender	ses	meanses	cses	mathach	size	sector	pracad	disclim	himinty	meansesBYcses	sectorBYcses
1224	1	0	1	-1.53	-0.43	-1.10	5.88	842	0	0.35	1.60	0	0.47	0
1224	2	0	1	-0.59	-0.43	-0.16	19.71	842	0	0.35	1.60	0	0.07	0
1224	3	0	0	-0.53	-0.43	-0.10	20.35	842	0	0.35	1.60	0	0.04	0
1224	4	0	0	-0.67	-0.43	-0.24	8.78	842	0	0.35	1.60	0	0.10	0
1224	5	0	0	-0.16	-0.43	0.27	17.90	842	0	0.35	1.60	0	-0.12	0
1224	6	0	0	0.02	-0.43	0.45	4.58	842	0	0.35	1.60	0	-0.19	0
1224	7	0	1	-0.62	-0.43	-0.19	-2.83	842	0	0.35	1.60	0	0.08	0
1224	8	0	0	-1.00	-0.43	-0.57	0.52	842	0	0.35	1.60	0	0.24	0
1224	9	0	1	-0.89	-0.43	-0.46	1.53	842	0	0.35	1.60	0	0.20	0
1224	10	0	0	-0.46	-0.43	-0.03	21.52	842	0	0.35	1.60	0	0.01	0
1224	11	0	1	-1.45	-0.43	-1.02	9.48	842	0	0.35	1.60	0	0.44	0
1224	12	0	1	-0.66	-0.43	-0.23	16.06	842	0	0.35	1.60	0	0.10	0
1224	13	0	0	-0.47	-0.43	-0.04	21.18	842	0	0.35	1.60	0	0.02	0
1224	14	0	1	-0.99	-0.43	-0.56	20.18	842	0	0.35	1.60	0	0.24	0
1224	15	0	0	0.33	-0.43	0.76	20.35	842	0	0.35	1.60	0	-0.33	0
1224	16	0	1	-0.68	-0.43	-0.25	20.51	842	0	0.35	1.60	0	0.11	0
1224	17	0	0	-0.30	-0.43	0.13	19.34	842	0	0.35	1.60	0	-0.06	0
1224	18	1	0	-1.53	-0.43	-1.10	4.14	842	0	0.35	1.60	0	0.47	0
1224	19	0	1	0.04	-0.43	0.47	2.93	842	0	0.35	1.60	0	-0.20	0
1224	20	0	0	-0.08	-0.43	0.35	16.41	842	0	0.35	1.60	0	-0.15	0
1224	21	0	1	0.06	-0.43	0.49	13.65	842	0	0.35	1.60	0	-0.21	0
1224	22	0	1	-0.13	-0.43	0.30	6.56	842	0	0.35	1.60	0	-0.13	0
1224	23	0	1	0.47	-0.43	0.90	9.65	842	0	0.35	1.60	0	-0.39	0

First few lines of `hsb.dta`

We take the standardized measure of mathematics achievement (`mathach`) as the student-level outcome, y_{ij}. The student level (level-1) explanatory variables are the student socio-economic status, $\texttt{ses}_{ij}$, which is a composite of parental education, occupation and income; an indicator for student `minority` ($1 =$ yes, $0 =$ other), and an indicator for student `gender` ($1 =$ female, $0 =$ male). There are two school-level (level-2) variables: a school-level variable `sector`, which is an indicator variable taking on a value of one for Catholic schools and zero for public schools, and an aggregate of school-level characteristics $(\texttt{meanses})_j$, the average of the student `ses` values within each school. Two variables `ses` and `meanses` are centred at the grand mean.

Questions motivating these analyses include the following:

- How much do the high schools vary in their mean mathematics achievement?
- Do schools with high `meanses` also have high maths achievement?
- Is the strength of association between student `ses` and `mathach` similar across schools?
- Is `ses` a more important predictor of achievement in some schools than in others?
- How do public and Catholic schools compare in terms of mean `mathach` and in terms of the strength of the `ses`- relationship, after we control for `meanses`?

To obtain some preliminary information about how much variation in the outcome lies within and between schools, we may fit the one-way ANOVA to the high school data.

The student-level model is

$$y_{ij} = \beta_{0j} + \varepsilon_{ij},$$

where y_{ij} is `mathach`, for $i = 1, \cdots, n_j$ students in school j, and $j = 1, \cdots, 160$ schools. At the school level (level 2), each school's mean maths achievement, β_{0j}, is represented as a function of the grand mean, γ_{00}, plus a random error, u_{0j}. We refer to the variance of u_{0j} as the school-level variance and to the variance of ε_{ij} as the student-level variance.

The combined model is given by

$$y_{ij} = \gamma_{00} + u_{0j} + \varepsilon_{ij}.$$

The data can be read into Sabre and this model estimated.

1.6 Including School-Level Effects - Model 2

The simple model $y_{ij} = \beta_{0j} + \varepsilon_{ij}$ provides a baseline against which we can compare more complex models. We begin with the inclusion of one level-2 variable, `meanses`, which indicates the average `ses` of children within each school. Each school's mean is now predicted by the `meanses` of the school:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\texttt{meanses}_j + u_{0j},$$

where γ_{00} is the intercept, γ_{01} is the effect of `meanses` on β_{0j}, and we assume $u_{0j} \sim N\left(0, \sigma^2_{u_0}\right)$. Substituting the level-2 equation into the level-1 model yields

$$y_{ij} = [\gamma_{00} + \gamma_{01}\texttt{meanses}_j] + [u_{0j} + \varepsilon_{ij}].$$

This model is the sum of two parts: a fixed part and a random part. The two terms in the first bracket represent the fixed part, consisting of the two gamma terms. The two terms in the second bracket represent the random part, consisting of the u_{0j} (which represents variation between schools) and the ε_{ij} (which represents variation within schools).

We note that the variance components $\sigma^2_{u_0}$ and σ^2_{ε} now have different meanings. In the model $y_{ij} = \beta_{0j} + \varepsilon_{ij}$, there were no explanatory variables, so $\sigma^2_{u_0}$ and σ^2_{ε} were unconditional components. Having added a predictor, $\sigma^2_{u_0}$ and σ^2_{ε} are now conditional components. The variance $\sigma^2_{u_0}$ is a residual or conditional variance, that is, $var\left(\beta_{0j}|\texttt{meanses}\right)$, the school-level variance in β_{0j} after controlling for school `meanses`.

1.6.1 Sabre commands

```
log using hsb1_s.log, replace
set more off
use hsb
#delimit ;
```

```
sabre, data school student minority gender ses meanses cses mathach size
       sector pracad disclim himinty meansesBYcses sectorBYcses;
sabre school student minority gender ses meanses cses mathach size sector
      pracad disclim himinty meansesBYcses sectorBYcses, read;
#delimit cr
sabre, case school
sabre, yvar mathach
sabre, family g
sabre, constant cons
sabre, lfit cons
sabre, dis m
sabre, dis e
sabre, mass 64
sabre, fit cons
sabre, dis m
sabre, dis e
sabre, lfit meanses cons
sabre, dis m
sabre, dis e
sabre, fit meanses cons
sabre, dis m
sabre, dis e
log close
clear
exit
```

The command `log using hsb1_s.log, replace` opens the file `hsb1.log` for the log file from the Sabre analysis and deletes any previous file with the same name. The command `sabre, lfit cons` estimates the homogeneous model using OLS, and the command `sabre, mass 64` is used to provide a good approximation to the integral in $L\left(\gamma_{00}, \sigma_{\varepsilon}^2, \sigma_{u_0}^2 | \mathbf{y}\right)$ for the linear model. Adaptive quadrature would require fewer mass points. An explanation of all the commands is to be found in the Sabre manual. The file `hsb1_s.log` would contain the following results.

1.6.2 Sabre log file

```
(Standard Homogenous Model)


    Parameter              Estimate         Std. Err.
    ___________________________________________________
    (intercept)             12.748          0.81145E-01
    sigma                   6.8782


(Random Effects Model)


    Parameter              Estimate         Std. Err.
    ___________________________________________________
    (intercept)             12.637          0.24359
    sigma                   6.2569          0.52794E-01
```

```
scale                  2.9246          0.18257

X-vars            Y-var            Case-var
------------------------------------------------
(intercept)       response         case.1

Univariate model
Standard linear
Gaussian random effects

Number of observations            =    7185
Number of cases                   =     160

X-var df           =     1
Sigma df           =     1
Scale df           =     1

Log likelihood =     -23557.905     on    7182 residual degrees of freedom
```

```
(Standard Homogenous Model)


    Parameter              Estimate         Std. Err.
    --------------------------------------------------
    (intercept)             12.713          0.76215E-01
    meanses                 5.7168          0.18429
    sigma                   6.4596
```

```
(Random Effects Model)


    Parameter              Estimate         Std. Err.
    --------------------------------------------------
    (intercept)             12.650          0.14834
    meanses                 5.8629          0.35917
    sigma                   6.2576          0.52800E-01
    scale                   1.6103          0.12314

    X-vars            Y-var            Case-var
    ------------------------------------------------
    (intercept)       response         case.1
    meanses

    Univariate model
    Standard linear
    Gaussian random effects

    Number of observations            =    7185
    Number of cases                   =     160

    X-var df           =     2
    Sigma df           =     1
    Scale df           =     1

    Log likelihood =     -23479.554     on    7181 residual degrees of freedom
```

1.6.3 Model 1 discussion

The estimate of the grand mean, γ_{00}, is 12.637. This mean should be interpreted as the expected value of the maths achievement for a random student in a randomly drawn class. The log file also shows that the estimate of the within-school variance component $(6.2569)^2 = 39.149$ is nearly five times the size of the between-school variance component $(2.9246)^2 = 8.5533$. These variance component estimates give an intraclass correlation coefficient estimate of $\hat{\rho} = 8.5533/(8.5533 + 39.149) = 0.179$ indicating that about 18% of the variance in maths achievement is between schools.

1.6.4 Model 2 discussion

The estimated regression equation is given by

$$y_{ij} = [12.650 + 5.8629\ \texttt{meanses}_j] + [u_{0j} + \varepsilon_{ij}].$$

The coefficient of `cons`, 12.65, estimates γ_{00}, the mean maths achievement when the remaining predictors (here, just `meanses`) are 0. Because `meanses` is centred at the grand mean, γ_{00} is the estimated `mathach` in a school of "average meanses" . The coefficient of `meanses`, 5.8629, provides our estimate of the other fixed effect, γ_{01}, and tells us about the relationship between maths achievement and `meanses`.

We note that the conditional component for the within-school variance (the residual component representing σ_ε^2) has remained virtually unchanged (going from $(6.2569)^2$ to $(6.2576)^2$). The variance component representing variation between schools, however, has diminished markedly (going from $(2.9246)^2$ to $(1.6103)^2$). This tells us that the predictor `meanses` explains a large proportion of the school-to-school variation in mean maths achievement.

The estimated ρ is now a conditional intraclass correlation coefficient and measures the degree of dependence among observations within schools after controlling for the effect of `meanses`. This conditional estimate of

$$\hat{\rho} = (1.6103)^2 / ((1.6103)^2 + (6.2576)^2) = 0.062,$$

which is much smaller than the unconditional one.

1.7 Exercises

There are also two exercises to accompany this material, namely C1 and L1.

1.8 References

Goldstein, H., (1987), Multilevel Models in Educational and Social Research, Griffin, London.

Hsiao, C., (1986), Analysis of Panel Data, Cambridge University Press, Cambridge.

Rabe-Hesketh, S., and Skrondal, A., (2005), Multilevel and Longitudinal Modelling using Stata, Stata Press, Stata Corp, College Station, Texas.

Wooldridge, J. M. (2006), Introductory Econometrics: A Modern Approach. Third edition. Thompson, Australia.

Chapter 2

Linear Models II

2.1 Introduction

The basic idea of multilevel analysis is that data sets with a nesting structure that includes unexplained variability at each level of nesting are usually not adequately represented by multiple regression. The reason is that the unexplained variability in single-level multiple regression analysis is only the variance of the residual term. Variability in multilevel data, however, has a more complicated structure related to the fact that several populations are involved in modelling such data: one population for each level. Explaining variability in a multilevel structure can be achieved by explaining variability between level-1 units but also by explaining variability between higher-level units. For example, in fitting multilevel models with two levels, we can try to explain the variability between level-2 units if a random intercept at level 2 exists.

2.2 Two-Level Random Intercept Models

In these models, the intercept β_{0j} does depend on the level-2 units but the regression coefficient of the x_{ij} is constant. The resulting model with one explanatory variable x_{ij} is given by

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \varepsilon_{ij}.$$

For the level-2 model, the group-dependent intercept can be split into an grand mean intercept and the group-dependent deviation:

$$\beta_{0j} = \gamma_{00} + u_{0j},$$

and the same fixed effect of x_{ij} for each level-2 unit is assumed:

$$\beta_{1j} = \gamma_{10}.$$

The grand mean is γ_{00} and the regression coefficient for x_{ij} is γ_{10}. Substitution now leads to the model

$$y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + u_{0j} + \varepsilon_{ij}.$$

The random effects u_{0j} are the level-2 unit residuals, controlling for the effects of variable x_{ij}. It is assumed that these residuals are drawn from normally distributed populations having zero mean and a constant variance $\sigma^2_{u_0}$, given the values x_{ij} of the explanatory variable. The population mean and variance of the level-1 unit residuals ε_{ij} are assumed to be zero and σ^2_ε, respectively across the level-2 units.

The variance of y_{ij} conditional on the value of x_{ij} is given by

$$var(y_{ij}|x_{ij}) = var(u_{0j}) + var(\varepsilon_{ij}) = \sigma^2_{u_0} + \sigma^2_\varepsilon,$$

while the covariance between two different level-1 units (i and i', with $i \neq i'$) in the same level-2 unit is

$$cov(y_{ij}, y_{i'j}|x_{ij}, x_{i'j}) = var(u_{0j}) = \sigma^2_{u_0}.$$

The fraction of residual variability that can be attributed to level one is given by

$$\frac{\sigma^2_\varepsilon}{\sigma^2_{u_0} + \sigma^2_\varepsilon},$$

and for level two this fraction is

$$\frac{\sigma^2_{u_0}}{\sigma^2_{u_0} + \sigma^2_\varepsilon}.$$

The residual intraclass correlation coefficient,

$$\rho\left(y_{ij}|x_{ij}\right) = \frac{\sigma^2_{u_0}}{\sigma^2_{u_0} + \sigma^2_\varepsilon},$$

is the correlation between the y-values of any two different level-1 units in the same level-2 unit, controlling for variable x. It is analogous to the usual intraclass correlation coefficient, but now controls for x. If the residual intraclass correlation coefficient, or equivalently, $\sigma^2_{u_0}$, is positive, then the hierarchical linear model is a better analysis than ordinary least squares regression.

An extension of this model allows for the introduction of level-2 predictors z_j. Using the level-2 model

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01} z_j + u_{0j}, \\ \beta_{1j} &= \gamma_{10}, \end{aligned}$$

the model becomes

$$y_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + \gamma_{01} z_j + u_{0j} + \varepsilon_{ij},$$

so that

$$\mu_{ij} = \gamma_{00} + \gamma_{10} x_{ij} + \gamma_{01} z_j + u_{0j}.$$

This model provides for a level-2 predictor, z_j, while also controlling for the effect of a level-1 predictor, x_{ij}, and the random effects of the level-2 units, u_{0j}.

2.3 General Two-Level Models Including Random Intercepts

Just as in multiple regression, more than one explanatory variable can be included in the random intercept model. When the explanatory variables at the individual level are denoted by $x_1, \cdots, x_P$, and those at the group level by $z_1, \cdots, z_Q$, adding their effects to the random intercept model leads to the following formula

$$y_{ij} = \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + u_{0j} + \varepsilon_{ij},$$

so that

$$\mu_{ij} = \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + u_{0j}.$$

The regression parameters γ_{p0} $(p = 1, \cdots, P)$ and γ_{0q} $(q = 1, \cdots, Q)$ for level-one and level-two explanatory variables, respectively, again have the same interpretation as regression coefficients in multiple regression models: one unit increase in the value of x_p (or z_q) is associated with an average increase in y of γ_{p0} (or γ_{0q}) units. Just as in multiple regression, some of the variables x_p and z_q may be interaction variables, or non-linear (e.g., quadratic) transforms of basic variables.

The first part of the right-hand side of the above equation incorporating the regression coefficients,

$$\gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj},$$

is called the fixed part of the model, because the coefficients are fixed (i.e., not stochastic). The remaining part,

$$u_{0j} + \varepsilon_{ij},$$

is called the random part of the model. It is again assumed that all residuals, u_{0j} and ε_{ij}, are mutually independent and have zero means conditional on the explanatory variables. A somewhat less crucial assumption is that these residuals are drawn from normally distributed populations. The population variance of the level-one residuals ε_{ij} is denoted by σ_ε^2 while the population variance of the level-two residuals u_{0j} is denoted by $\sigma_{u_0}^2$.

2.4 Likelihood

$$L\left(\gamma, \sigma_\varepsilon^2, \sigma_{u_0}^2 | \mathbf{y}, \mathbf{x}, \mathbf{z}\right) = \prod_j \int_{-\infty}^{+\infty} \prod_i g\left(\mathbf{y}_{ij} | \mathbf{x}_{ij}, \mathbf{z}_j, u_{0j}\right) f\left(u_{0j}\right) du_{0j},$$

where

$$g\left(\mathbf{y}_{ij}|\mathbf{x}_{ij},\mathbf{z}_j,u_{0j}\right)=\frac{1}{\sqrt{2\pi}\sigma_\varepsilon}\exp\left(-\frac{\left[y_{ij}-\mu_{ij}\right]^2}{2\sigma_\varepsilon^2}\right),$$

$$\mu_{ij}=\gamma_{00}+\sum_{p=1}^{P}\gamma_{p0}x_{pij}+\sum_{q=1}^{Q}\gamma_{0q}z_{qj}+u_{0j},$$

and

$$f\left(u_{0j}\right)=\frac{1}{\sqrt{2\pi}\sigma_{u_0}}\exp\left(-\frac{u_{0j}^2}{2\sigma_{u_0}^2}\right).$$

2.5 Residuals

In a single-level model the usual estimate of the single residual term is just the residual

$$e_{ij}=y_{ij}-\hat{\gamma}_{00}-\hat{\gamma}_{10}x_{ij}.$$

In a multilevel model, however, there are several residuals at different levels. In a random intercept model, the level-2 residual u_{0j} can be predicted by the posterior means

$$\widehat{u}_{0j}=E\left(u_{0j}|\mathbf{y}_j,\mathbf{x}_j,\theta\right),$$

where θ are the model parameters. We can show that

$$\widehat{u}_{0j}=\frac{\sigma_{u_0}^2}{\sigma_{u_0}^2+\sigma_\varepsilon^2/n_j}\overline{e}_j,$$

where the $\overline{e}_j$ are averages of e_{ij} for level-2 units $j=1,\cdots,N$. These residuals have two interpretations. Their basic interpretation is as random variables with a distribution whose parameter values tell us about the variation among the level-2 units, and which provide efficient estimates for the fixed coefficients. A second interpretation is as individual estimates for each level-2 unit where we use the assumption that they belong to a population of units to predict their values.

When the residuals at higher levels are of interest in their own right, we need to be able to provide interval estimates and point estimates for them. For these purposes, we require estimates of the standard errors of the estimated residuals, where the sample estimate is viewed as a random realization from repeated sampling of the same higher-level units whose unknown true values are of interest.

Note that we can now estimate the level-1 residuals simply by the formula:

$$\hat{\varepsilon}_{ij}=e_{ij}-\widehat{u}_{0j}.$$

The level-1 residuals are generally not of interest in their own right but are used rather for model checking, having first been standardised using the diagnostic standard errors.

2.6 Checking Assumptions in Multilevel Models

Residual plots can be used to check model assumptions. There is one important difference from ordinary regression analysis; there is more than one residual. In fact, we have residuals for each random effect in the multilevel model. Consequently, many different residual plots can be constructed.

Most regression assumptions are concerned with residuals; the difference between the observed y and the y predicted by the regression line. These residuals will be very useful to test whether or not the multilevel model assumptions hold.

As in single-level models, we can use the estimated residuals to help check the model assumptions. The two particular assumptions that can be studied readily are the assumption of normality and the assumption that the variances in the model are constant. Because the variances of the residual estimates depend in general on the values of the fixed coefficients it is common to standardise the residuals by dividing by the appropriate standard errors.

To examine the assumption of linearity, for example, we can produce a residual plot against predicted values of the dependent variable using the fixed part of the multilevel regression model for the prediction. A residual plot should show a random scatter of residuals around the zero line. Even if the residuals are evenly distributed around zero, the regression model is still questionable when there is a pattern in the residuals. Ideally, you should not be able to detect any patterns.

To check the normality assumption we can use a normal probability plot. The standardized residuals are plotted against a theoretical normal distribution in such a way that the points should form an approximate straight line. Departures from this straight line indicate departures from normality. We will return to residuals in a later section, though Sabre doesnt currently make the residuals available to Stata.

2.7 Example C2. Linear model of Pupil's Maths Achievement

The data we use in this example (`hsb.dta`) are a sub-sample from the 1982 High School and Beyond Survey (Raudenbush and Bryk, 2002), and include information on 7,185 students nested within 160 schools: 90 public and 70 Catholic. Sample sizes vary from 14 to 67 students per school.

2.7.1 References

Raudenbush, S.W., Bryk, A.S., 2002, Hierarchical Linear Models, Thousand Oaks, CA. Sage.

2.7.2 Data description for `hsb.dta`

Number of observations (rows): 7185
Number of level-2 cases: 160

2.7.3 Variables

The variables include the following:
`school`: school identifier
`student`: student identifier
`minority`: 1 if student is from an ethnic minority, 0 if otherwise)
`gender`: 1 if student is female, 0 otherwise
`ses`: a standardized scale constructed from variables measuring parental education, occupation, income and socio-economic status
`meanses`: mean of the SES values for the students in this school
`mathach`: a measure of the students' mathematics achievement
`size`: school enrolment
`sector`: 1 if school is from the Catholic sector, 0 if public
`pracad`: proportion of students in the academic track
`disclim`: a scale measuring disciplinary climate
`himnty`: 1 if more than 40% minority enrolment, 0 if less than 40%

school	student	minority	gender	ses	meanses	cses	mathach	size	sector	pracad	disclim	himinty	meansesBYcses	sectorBYcses
1224	1	0	1	-1.53	-0.43	-1.10	5.88	842	0	0.35	1.60	0	0.47	0
1224	2	0	1	-0.59	-0.43	-0.16	19.71	842	0	0.35	1.60	0	0.07	0
1224	3	0	0	-0.53	-0.43	-0.10	20.35	842	0	0.35	1.60	0	0.04	0
1224	4	0	0	-0.67	-0.43	-0.24	8.78	842	0	0.35	1.60	0	0.10	0
1224	5	0	0	-0.16	-0.43	0.27	17.90	842	0	0.35	1.60	0	-0.12	0
1224	6	0	0	0.02	-0.43	0.45	4.58	842	0	0.35	1.60	0	-0.19	0
1224	7	0	1	-0.62	-0.43	-0.19	-2.83	842	0	0.35	1.60	0	0.08	0
1224	8	0	0	-1.00	-0.43	-0.57	0.52	842	0	0.35	1.60	0	0.24	0
1224	9	0	1	-0.89	-0.43	-0.46	1.53	842	0	0.35	1.60	0	0.20	0
1224	10	0	0	-0.46	-0.43	-0.03	21.52	842	0	0.35	1.60	0	0.01	0
1224	11	0	1	-1.45	-0.43	-1.02	9.48	842	0	0.35	1.60	0	0.44	0
1224	12	0	1	-0.66	-0.43	-0.23	16.06	842	0	0.35	1.60	0	0.10	0
1224	13	0	0	-0.47	-0.43	-0.04	21.18	842	0	0.35	1.60	0	0.02	0
1224	14	0	1	-0.99	-0.43	-0.56	20.18	842	0	0.35	1.60	0	0.24	0
1224	15	0	0	0.33	-0.43	0.76	20.35	842	0	0.35	1.60	0	-0.33	0
1224	16	0	1	-0.68	-0.43	-0.25	20.51	842	0	0.35	1.60	0	0.11	0
1224	17	0	0	-0.30	-0.43	0.13	19.34	842	0	0.35	1.60	0	-0.06	0
1224	18	1	0	-1.53	-0.43	-1.10	4.14	842	0	0.35	1.60	0	0.47	0
1224	19	0	1	0.04	-0.43	0.47	2.93	842	0	0.35	1.60	0	-0.20	0
1224	20	0	0	-0.08	-0.43	0.35	16.41	842	0	0.35	1.60	0	-0.15	0
1224	21	0	1	0.06	-0.43	0.49	13.65	842	0	0.35	1.60	0	-0.21	0
1224	22	0	1	-0.13	-0.43	0.30	6.56	842	0	0.35	1.60	0	-0.13	0
1224	23	0	1	0.47	-0.43	0.90	9.65	842	0	0.35	1.60	0	-0.39	0

First few lines of `hsb.dta`

We will use these data as a worked example. We think of our data as structured in two levels: students within schools and between schools. The outcome considered here is again maths achievement score (y) related to a set of explanatory variables x and z. At the student level,

$$y_{ij} = \beta_{0j} + \beta_{1j}\texttt{ses}_{ij} + \beta_{2j}\texttt{minority}_{ij} + \beta_{3j}\texttt{gender}_{ij} + \varepsilon_{ij}.$$

At the school level,

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\texttt{meanses}_j + u_{0j},$$

where $u_{0j} \sim N\left(0, \sigma^2_{u_0}\right)$, and

$$\beta_{pj} = \gamma_{p0}, \text{ for } p = 1, 2, 3.$$

In the combined form, the model is

$$\begin{aligned} y_{ij} = {} & \gamma_{00} + \gamma_{01}\texttt{meanses}_j + \gamma_{10}\texttt{ses}_{ij} + \gamma_{20}\texttt{minority}_{ij} \\ & + \gamma_{30}\texttt{gender}_{ij} + u_{0j} + \varepsilon_{ij}. \end{aligned}$$

Having written down a combined equation, we can now fit the model using Sabre.

2.7.4 Sabre commands

```
log using hsb2_s.log, replace
set more off
use hsb
#delimit ;
sabre, data school student minority gender ses meanses cses mathach size
       sector pracad disclim himinty meansesBYcses sectorBYcses;
sabre school student minority gender ses meanses cses mathach size sector
      pracad disclim himinty meansesBYcses sectorBYcses, read;
#delimit cr
```

```
sabre, case school
sabre, yvar mathach
sabre, family g
sabre, constant cons
sabre, lfit minority gender ses meanses cons
sabre, dis m
sabre, dis e
sabre, mass 64
sabre, fit minority gender ses meanses cons
sabre, dis m
sabre, dis e
log close
clear
exit
```

2.7.5 Sabre log file

```
(Standard Homogenous Model)

    Parameter              Estimate         Std. Err.
    ___________________________________________________
    (intercept)             14.070          0.11710
    minority               -2.3410          0.17381
    gender                 -1.3200          0.14658
    ses                     1.9551          0.11151
    meanses                 2.8675          0.21311
    sigma                   6.1857


(Random Effects Model)

    Parameter              Estimate         Std. Err.
    ___________________________________________________
    (intercept)             14.048          0.17491
    minority               -2.7282          0.20412
    gender                 -1.2185          0.16082
    ses                     1.9265          0.10844
    meanses                 2.8820          0.36521
    sigma                   5.9905          0.50554E-01
    scale                   1.5480          0.11885

    X-vars           Y-var            Case-var
    ________________________________________________
    (intercept)      response         case.1
    minority
    gender
    ses
    meanses

    Univariate model
    Standard linear
    Gaussian random effects

    Number of observations            =    7185
```

```
Number of cases                    =     160

X-var df           =     5
Sigma df           =     1
Scale df           =     1

Log likelihood =     -23166.634     on    7178 residual degrees of freedom
```

2.7.6 Discussion

These results show that the covariates in the model for `mathach` generally have larger standard errors in the random effects model than they do in the homogeneous model. These results also show that in the random effects model, the random effect scale parameter estimate is highly significant with a value 1.5480 (s.e. 0.11885), suggesting that students in the same school have correlated responses. Furthermore, students that are from an ethnic minority do worse than those who are not, and female students seem to do worse than males.

For further material on the linear model with random intercepts see: Goldstein, (1987), Hsiao, (1986), Rabe-Hesketh and Skrondal (2005) and Wooldridge, J. M. (2006),

2.8 Comparing Model Likelihoods

Each model that is fitted to the same set of data has a corresponding log-likelihood value that is calculated at the maximum likelihood estimates for that model. These values are used to compare and statistically test terms in the model.

The deviance test, or likelihood ratio test, is a quite general principle for statistical testing. In applications of the hierarchical linear model, this test is used mainly for multi-parameter tests and for tests about the fixed part as well as the random part of the model. The general principle is as follows.

When parameters of a statistical model are estimated by the maximum likelihood (ML) method, the estimation also provides the likelihood, which can be transformed into the deviance defined as minus twice the natural logarithm of the likelihood. This deviance can be regarded as a measure of lack of fit between model and data, but (in most statistical models) one cannot interpret the deviance directly, but only differences in deviance for several models fitted to the same data set.

In general, suppose that model one has t parameters, while model two is a subset of model one with only r of the t parameters so that $r < t$. Model one will have a higher log-likelihood than model two. For large sample sizes, the difference between these two likelihoods, when multiplied by two, will behave like the chi-square distribution with $(t - r)$ degrees of freedom. This can be used to test the null hypothesis that the $(t - r)$ parameters that are not in both models are zero. Sabre computes the log-likelihoods $log(L)$ (which are negative values). These values can be used directly to calculate the differences for statistical tests. Differences between nested likelihoods are called deviances, where:

$$D = -2[log(L_r) - log(L_t)],$$

$log(L_t)$ is the log likelihood for the extended model, and $log(L_r)$ is the log likelihood for the simpler model. With large sample sizes, D approximately follows a chi-square distribution with $(t - r)$ degrees of freedom.

For regression models we are estimating, the homogeneous model log likelihood = -23285.328 on 7179 residual degrees of freedom when compared to the random effects model log likelihood = -23166.634 on 7178 residual degrees of freedom, here has a χ^2 improvement of $-2(-23285.328 + 23166.634) = 237.39$ for 1 df, which is highly significant, justifying the extra scale parameter.

The estimates of the residual variance σ_ε^2 and the random intercept variance $\sigma_{u_0}^2$ are much lower in the random effects model than in the simple model with no explanatory variables. This shows that a part of the variability is explained by including the explanatory variables at both levels. The residual intraclass correlation coefficient is estimated by

$$\hat{\rho} = \frac{(1.5480)^2}{(1.5480)^2 + (5.9905)^2} = 0.062595,$$

In a model without the explanatory variables, this was 0.18. The residual (or between-student) variation clearly dominates this model. The explanatory variables will have accounted for a good deal of the level-2 variance.

2.9 Exercises

There are also two exercises to accompany this section, namely C2 and L2.

2.10 References

Goldstein, H., (1987), Multilevel Models in Educational and Social Research, Griffin, London.

Hsiao, C., (1986), Analysis of Panel Data, Cambridge University Press, Cambridge.

Rabe-Hesketh, S., and Skrondal, A., (2005), Multilevel and Longitudinal Modelling using Stata, Stata Press, Stata Corp, College Station, Texas.

Wooldridge, J. M. (2006), Introductory Econometrics: A Modern Approach. Third edition. Thompson, Australia.

Chapter 3

Multilevel Binary Response Models

3.1 Introduction

In all of the multilevel linear models considered so far, it was assumed that the response variable has a continuous distribution and that the random coefficients and residuals are normally distributed. These models are appropriate where the expected value of the response variable at each level may be represented as a linear function of the explanatory variables. The linearity and normality assumptions can be checked using standard graphical procedures. There are other kinds of outcomes, however, for which these assumptions are clearly not realistic. An example is the model for which the response variable is discrete.

Important instances of discrete response variables are binary variables (e.g., success vs. failure of whatever kind) and counts (e.g., in the study of some kind of event, the number of events happening in a predetermined time period).

For a binary variable y_{ij} that has probability μ_{ij} for outcome 1 and probability $1 - \mu_{ij}$ for outcome 0, the mean is

$$E(y_{ij}) = \mu_{ij},$$

and the variance is

$$var(y_{ij}) = \mu_{ij}(1 - \mu_{ij}).$$

The variance is not a free parameter but is determined by the mean.

This has led to the development of regression-like models that differ from the usual multiple linear regression models and that take account of the non-normal distribution of the response variable, its restricted range, and the relation between mean and variance. The best-known method of this kind is logistic regression, a regression-like model for binary data.

3.2 The Two-Level Logistic Model

We start by introducing a simple two-level model that will be used to illustrate the analysis of binary response data. Let j denote the level-2 units (clusters) and i denote the level-1 units (nested observations). Assume that there are $j = 1, \cdots, m$ level-2 units and $i = 1, \cdots, n_j$ level-1 units nested within each level-2 unit j. The total number of level-1 observations across level-2 units is given by $n = \sum_{j=1}^{m} n_j$.

For a multilevel representation of a simple model with only one explanatory variable x_{ij}, the level-1 model is written in terms of the latent response variable y_{ij}^* as

$$y_{ij}^* = \beta_{0j} + \beta_{1j} x_{ij} + \varepsilon_{ij},$$

and the level-2 model becomes

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + u_{0j}, \\ \beta_{1j} &= \gamma_{10}. \end{aligned}$$

In practice, y_{ij}^* is unobservable, and this can be measured indirectly by an observable binary variable y_{ij} defined by

$$y_{ij} = \begin{cases} 1 & if \quad y_{ij}^* > 0 \\ 0 & otherwise, \end{cases}$$

such that,

$$\begin{aligned} \Pr(y_{ij} = 1 \mid x_{ij}, u_{0j}) &= \Pr\left(y_{ij}^* > 0 \mid u_{0j}\right) \\ &= \Pr(\gamma_{00} + \gamma_{10} x_{ij} + u_{0j} + \varepsilon_{ij} > 0 \mid u_{0j}) \\ &= \Pr(\varepsilon_{ij} > -\{\gamma_{00} + \gamma_{10} x_{ij} + u_{0j}\} \mid u_{0j}) \\ &= \int_{-\{\gamma_{00} + \gamma_{10} x_{ij} + u_{0j}\}}^{\infty} f(\varepsilon_{ij} \mid u_{0j})\, d\varepsilon_{ij} \\ &= 1 - F(-\{\gamma_{00} + \gamma_{10} x_{ij} + u_{0j}\}) \\ &= \mu_{ij}. \end{aligned}$$

For symmetric distributions for $f(\varepsilon_{ij} \mid u_{0j})$ like the normal or logistic we have

$$1 - F(-\{\gamma_{00} + \gamma_{10} x_{ij} + u_{0j}\}) = F(\gamma_{00} + \gamma_{10} x_{ij} + u_{0j}),$$

where $F(\cdot)$ is the cumulative distribution function of ε_{ij}.

We view the observed values y_{ij} as a realization of a random variable y_{ij} that can take the values one and zero with probabilities μ_{ij} and $1 - \mu_{ij}$, respectively. The distribution of y_{ij} is called a Bernoulli distribution with parameter μ_{ij}, and can be written as

$$g(y_{ij} | x_{ij}, u_{0j}) = \mu_{ij}^{y_{ij}} (1 - \mu_{ij})^{1 - y_{ij}}, \quad y_{ij} = 0, 1.$$

To proceed, we need to impose an assumption about the distributions of u_{0j} and ε_{ij}. As in the linear case, we assume that the u_{0j} is distributed as $N\left(0, \sigma^2_{u_0}\right)$. Then, if the cumulative distribution of ε_{ij} is assumed to be logistic, we have the multilevel logit model, and if we assume that $\varepsilon_{ij} \sim N(0,1)$, we have the probit model.

We complete the specification of the logit model by expressing the functional form for μ_{ij} in the following manner:

$$\mu_{ij} = \frac{exp\left(\gamma_{00} + \gamma_{10}x_{ij} + u_{0j}\right)}{1 + exp\left(\gamma_{00} + \gamma_{10}x_{ij} + u_{0j}\right)}.$$

The probit model is based upon the assumption that the disturbances ε_{ij} are independent standard normal variates, such that

$$\mu_{ij} = \Phi(\gamma_{00} + \gamma_{10}x_{ij} + u_{0j}),$$

where $\Phi(\cdot)$ denotes the cumulative distribution function for a standard normal variable.

3.3 Logit and Probit Transformations

Interpretation of the parameter estimates obtained from either the logit or probit regressions are best achieved on a linear scale, such that for a logit regression, we can re-express μ_{ij} as

$$logit\left(\mu_{ij}\right) = log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \gamma_{00} + \gamma_{10}x_{ij} + u_{0j}.$$

This equation represents the log odds of observing the response $y_{ij} = 1$. This is linear in x, and so the effect of a unit change in x_{ij} is to increase the log odds by γ_{10}. Because the logit link function is non-linear, the effect of a unit increase in x_{ij} is harder to comprehend if measured on the probability scale μ_{ij}.

The probit model may be rewritten as

$$probit\left(\mu_{ij}\right) = \Phi^{-1}\left(\mu_{ij}\right) = \gamma_{00} + \gamma_{10}x_{ij} + u_{0j}.$$

The logistic and normal distributions are both symmetrical around zero and have very similar shapes, except that the logistic distribution has fatter tails. As a result, the conditional probability functions are very similar for both models, except in the extreme tails. For both the logit and probit link functions, any probability value in the range $[0,1]$ is transformed so that the resulting values of $logit(\mu_{ij})$ and $probit(\mu_{ij})$ will lie between $-\infty$ and $+\infty$.

A further transformation of the probability scale that is sometimes useful in modelling binomial data is the complementary log-log transformation. This function again transforms a probability μ_{ij} in the range $[0,1]$ to a value in $(-\infty, +\infty)$, using the relationship $log[-log(1-\mu_{ij})]$.

3.4 General Two-Level Logistic Models

Suppose the observed binary responses are binomially distributed, such that $y_{ij} \sim bin\,(1, \mu_{ij})$, with conditional variance $var(y_{ij}|\mu_{ij}) = \mu_{ij}\,(1 - \mu_{ij})$. The multilevel logistic regression model with P level-1 explanatory variables $x_1, \cdots, x_P$ and Q level-2 explanatory variables $z_1, \cdots, z_Q$ has the following form:

$$logit(\mu_{ij}) = \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + u_{0j},$$

where it is assumed that u_{0j} has a normal distribution with zero mean and variance $\sigma^2_{u_0}$.

3.5 Residual Intraclass Correlation Coefficient

For binary responses, the intraclass correlation coefficient is often expressed in terms of the correlation between the latent responses y^*. Since the logistic distribution for the level-1 residual, ε_{ij}, implies a variance of $\pi^2/3 = 3.29$, this implies that for a two-level logistic random intercept model with an intercept variance of $\sigma^2_{u_0}$, the intraclass correlation coefficient is

$$\rho = \frac{\sigma^2_{u_0}}{\sigma^2_{u_0} + \pi^2/3}.$$

For a two-level random intercept probit model, this type of intraclass correlation coefficient becomes

$$\rho = \frac{\sigma^2_{u_0}}{\sigma^2_{u_0} + 1},$$

since for the probit model we assume that $\varepsilon_{ij} \sim N\,(0, 1)$, and this model fixes the level-1 residual variance of the unobservable variable y^* to 1 (see, e.g., Skrondal and Rabe-Hesketh, 2004).

3.6 Likelihood

$$L\left(\gamma, \sigma^2_{u_0}|\mathbf{y}, \mathbf{x}, \mathbf{z}\right) = \prod_j \int_{-\infty}^{+\infty} \prod_i g\left(\mathbf{y}_{ij}|\mathbf{x}_{ij}, \mathbf{z}_j, u_{0j}\right) f\left(u_{0j}\right) du_{0j},$$

where

$$g\left(\mathbf{y}_{ij}|\mathbf{x}_{ij}, \mathbf{z}_j, u_{0j}\right) = \mu_{ij}^{y_{ij}} \left(1 - \mu_{ij}\right)^{1-y_{ij}},$$

$$\mu_{ij} = 1 - F\left(-\left\{\gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + u_{0j}\right\}\right),$$

and

$$f(u_{0j}) = \frac{1}{\sqrt{2\pi}\sigma_{u_0}} \exp\left(-\frac{u_{0j}^2}{2\sigma_{u_0}^2}\right).$$

Sabre evaluates the integral $L\left(\gamma, \sigma_{u_0}^2 | \mathbf{y}, \mathbf{x}, \mathbf{z}\right)$ for the binary response model using normal Gaussian quadrature or adaptive Gaussian quadrature (numerical integration). There is not an analytic solution for this integral with normally distributed u_{0j}.

A cross-sectional example will be demonstrated.

3.7 Example C3. Binary Response Model of Pupil's Repeating a Grade at Primary School

Raudenbush and Bhumirat (1992) analysed data on whether not 7185 children had to repeat a grade during their time at primary school (we used the data from 411 schools). The data were from a national survey of primary education in Thailand in 1988. We use a subset of the Raudenbush and Bhumirat (1992) data.

3.7.1 References

Raudenbush, S.W., Bhumirat, C., 1992. The distribution of resources for primary education and its consequences for educational achievement in Thailand, International Journal of Educational Research, 17, 143-164

3.7.2 Data description for `thaieduc1.dta`

Number of observations (rows): 8582
Number of level-2 cases: 411

3.7.3 Variables

`schoolid`: school identifier
`sex`: 1 if child is male, 0 otherwise
`pped`: 1 if the child had pre-primary experience, 0 otherwise
`repeat`: 1 if the child repeated a grade during primary school, 0 otherwise

schoolid	sex	pped	repeat
10101	0	1	0
10101	0	1	0
10101	0	1	0
10101	0	1	0
10101	0	1	0
10101	0	1	0
10101	0	1	0
10101	0	1	0
10101	0	1	0
10101	0	1	0
10101	0	1	0
10101	0	1	0
10101	0	1	0
10101	0	1	0
10101	0	1	0
10101	1	1	0
10101	1	1	0
10101	1	1	0
10101	1	1	0
10102	0	0	0
10102	0	1	0
10102	0	1	0
10102	0	1	0
10102	0	1	0
10102	0	1	0
10102	0	1	0

First few lines of `thaieduc1.dta`

A second version of these data `thaieduc2.dta`, number of observations (rows): 7516, contains the name set of variables as `thaieduc1.dta` with the addition of one further variable, a school-level variable `msesc` where:

`msesc`: mean socio-economic staus score

We take `repeat` as the binary response variable, the indicator of whether a child has ever repeated a grade ($0 = no$, $1 = yes$). The level-1 explanatory variables are `sex` ($0 = girl$, $1 = boy$) and child pre-primary education `pped` ($0 = no$, $1 = yes$). The probability that a child will repeat a grade during the primary years, μ_{ij}, is of interest.

At first, we estimate a multi-level model with just a multilevel constant term and the school-specific random effect:

$$logit\,(\mu_{ij}) = \gamma_{00} + u_{0j},$$

where $u_{0j} \sim N\left(0, \sigma^2_{u_0}\right)$. This will allow us to determine the magnitude of variation between schools in grade repetition. Then we estimate a multilevel model which includes the school-level variable `msesc` and the child-level variables `sex` and `pped`.

$$logit\,(\mu_{ij}) = \gamma_{00} + \gamma_{10}\texttt{msesc}_{ij} + \gamma_{20}\texttt{sex}_{ij} + \gamma_{30}\texttt{pped}_{ij} + u_{0j}.$$

3.7.4 Sabre commands

```
log using thaieduc_s.log, replace
set more off
use thaieduc1
sabre, data schoolid sex pped repeat
sabre schoolid sex pped repeat, read
sabre, case schoolid
sabre, yvar repeat
sabre, constant cons
sabre, lfit cons
sabre, dis m
sabre, dis e
sabre, fit cons
sabre, dis m
sabre, dis e
clear
use thaieduc2
sabre, data schoolid sex pped repeat msesc
sabre schoolid sex pped repeat msesc, read
sabre, case schoolid
sabre, yvar repeat
sabre, constant cons
sabre, lfit msesc sex pped cons
sabre, dis m
sabre, dis e
sabre, fit msesc sex pped cons
sabre, dis m
sabre, dis e
log close
clear
exit
```

3.7.5 Sabre log file

```
(Standard Homogenous Model)

    Parameter              Estimate         Std. Err.
    ___________________________________________________
    (intercept)            -1.7738          0.30651E-01


(Random Effects Model)

    Parameter              Estimate         Std. Err.
    ___________________________________________________
    (intercept)            -2.1263          0.79655E-01
    scale                   1.2984          0.84165E-01

    X-vars           Y-var            Case-var
    ________________________________________________
    (intercept)      response         case.1
```

```
Univariate model
Standard logit
Gaussian random effects

Number of observations            =    8582
Number of cases                   =     411

X-var df          =     1
Scale df          =     1

Log likelihood =     -3217.2642     on    8580 residual degrees of freedom
```

(Standard Homogenous Model)

```
Parameter             Estimate        Std. Err.
___________________________________________________
(intercept)           -1.7832         0.58777E-01
msesc                -0.24149         0.93750E-01
sex                   0.42777         0.67637E-01
pped                 -0.56885         0.70421E-01
```

(Random Effects Model)

```
Parameter             Estimate        Std. Err.
___________________________________________________
(intercept)           -2.2280         0.10461
msesc                -0.41369         0.22463
sex                   0.53177         0.75805E-01
pped                 -0.64022         0.98885E-01
scale                  1.3026         0.72601E-01

X-vars            Y-var            Case-var
________________________________________________
(intercept)       response         case.1
msesc
sex
pped

Univariate model
Standard logit
Gaussian random effects

Number of observations            =    7516
Number of cases                   =     356

X-var df          =     4
Scale df          =     1

Log likelihood =     -2720.7581     on    7511 residual degrees of freedom
```

3.7.6 Discussion

For the constant-only model, the estimated average log-odds of repetition across primary schools, γ_{00}, is -2.1263, and the variance between schools in school-average log-odds of repetition, $\sigma^2_{u_0}$, is $(1.2984)^2 = 1.6858$.

The estimate of the residual intraclass correlation coefficient is given by

$$\hat{\rho} = \frac{1.6858}{(1.6858 + \pi^2/3)} = 0.33881.$$

The second data set `thaieduc2.tab` has fewer cases than the first `thaieduc1.tab` because of missing values on the additional school-level covariate, `msesc`. The variance between schools in `thaieduc2.tab` for the logit model with `msesc`, `sex` and `pped` , $\sigma^2_{u_0}$, is $(1.3026)^2 = 1.6968$, which is highly significant and the estimate of the residual intraclass correlation coefficient is

$$\hat{\rho} = \frac{1.6968}{1.6968 + \pi^2/3} = 0.34027.$$

As `sex` is a dummy variable indicating whether the pupil is a girl or a boy, it can be helpful to write down a pair of fitted models, one for each gender. By substituting the values 1 for boy and 0 for girl in `sex`, we get the boy's constant $-2.2280 + 0.53177 = -1.6962$, and we can write:

$$logit\left(\mu_{ij}; girl\right) = -2.2280 - 0.4137\texttt{msesc}_j - 0.64022\texttt{pped}_{ij} + u_{0j},$$

$$logit\left(\mu_{ij}; boy\right) = -1.6962 - 0.4137\texttt{msesc}_j - 0.64022\texttt{pped}_{ij} + u_{0j}.$$

The intercepts in these two models are quite different.

For further discussion on binary response models with random intercepts see: Hsiao (1986), Rabe-Hesketh and Skrondal (2005) and Wooldridge (2002).

3.8 Exercises

There are two exercises to accompany this section, namely C3 and L4.

3.9 References

Rabe-Hesketh, S., and Skrondal, A., (2005), Multilevel and Longitudinal Modelling using Stata, Stata Press, Stata Corp, College Station, Texas.

Hsiao, C., (1986), Analysis of Panel Data, Cambridge University Press, Cambridge.

Wooldridge, J. M. (2002), Econometric Analysis of Cross Section and Panel Data, MIT Press, Cambridge Mass.

Chapter 4

Multilevel Models for Ordered Categorical Variables

4.1 Introduction

Variables that have as outcomes a small number of ordered categories are quite common in the social and biomedical sciences. Examples of such variables are responses to questionnaire items (with outcomes, e.g., 'completely disagree', 'disagree', 'agree', 'completely agree'), and a test scored by a teacher as 'fail', 'satisfactory', or 'good', etc. Very useful models for this type of data are the multilevel ordered logistic regression model, also called the multilevel ordered logit model or the multilevel proportional odds model; and the closely related multilevel ordered probit model. This section is about multilevel models where the response variable is such an ordinal categorical variable.

When the number of categories is two, the dependent variable is binary. When the number of categories is rather large (10 or more), it may be possible to approximate the distribution by a normal distribution and apply the hierarchical linear model for continuous outcomes. The main issue in such a case is the homoscedasticity assumption: is it reasonable to assume that the variances of the random terms in the hierarchical linear model are constant? (The random terms in a random intercept model are the level-one residuals, ε_{ij}, and the random intercept, u_{0j}.) To check this, it is useful to investigate the skewness of the distribution. If in some groups, or for some values of the explanatory variables, the response variable follows distributions that are very skewed toward the lower or upper end of the scale, then the homoscedasticity assumption is likely to be violated.

If the number of categories is small (3 or 4), or if it is between 5 and, say, 10, and the distribution cannot be well approximated by a normal distribution, then statistical methods for ordered categorical outcomes can be useful.

It is usual to assign numerical values to the ordered categories, remembering that the values are arbitrary. We consider the values for the ordered categories are defined as $1, \cdots, C$, where C is the number of categories. Thus, on the four-point scale mentioned above, 'completely disagree' would get the value 1, 'disagree' would be represented by 2, 'agree' by 3, and 'completely agree' by the value 4. Let the C ordered response categories be coded as $c = 1, 2, \cdots, C$.

The multilevel ordered models can also be formulated as threshold models. The real line is divided by thresholds into C intervals, corresponding to the C ordered categories. The first threshold is γ_1. Threshold γ_1 defines the upper bound of the interval corrsponding to observed outcome 1. Similarly, threshold γ_{-1} defines the lower bound of the interval corrsponding to observed outcome C. Threshold γ_c defines the boundary between the intervals corresponding to observed outcomes $c-1$ and c (for $c = 2, \cdots, C-1$). The latent response variable is denoted by y^*_{ij} and the observed categorical variable y_{ij} is related to y^*_{ij} by the 'threshold model' defined as

$$y_{ij} = \begin{cases} 1 & if \quad -\infty < y^*_{ij} \le \gamma_1 \\ 2 & if \quad \gamma_1 < y^*_{ij} \le \gamma_2 \\ \vdots & \qquad \vdots \\ C & if \quad \gamma_{C-1} < y^*_{ij} < +\infty. \end{cases}$$

4.2 The Two-Level Ordered Logit Model

Consider the latent response variable y^*_{ij} for level-one unit i in level-two unit j and the observed categorical variable y_{ij} related to y^*_{ij}. The ordinal models can be written in terms of y^*_{ij}

$$y^*_{ij} = \theta_{ij} + \varepsilon_{ij},$$

where

$$\theta_{ij} = \beta_{0j} + \sum_{p=1}^{P} \beta_p x_{pij}.$$

In the absence of explanatory variables and random intercepts, the response variable y_{ij} takes on the values of c with probability

$$p_{ij(c)} = Pr(y_{ij} = c),$$

for $c = 1, \cdots, C$. As ordinal response models often utilize cumulative comparisons of the ordinal outcome, define the cumulative response probabilities for the C categories of the ordinal outcome y_{ij} as

$$P_{ij(c)} = Pr(y_{ij} \leq c) = \sum_{k=1}^{c} p_{ij(k)}, \quad c = 1, \cdots, C.$$

Note that this cumulative probability for the last category is 1; i.e. $P_{ij(C)} = 1$. Therefore, there are only $(C-1)$ cumulative probabilities $P_{ij(c)}$ to estimate. If the cumulative density function of ε_{ij} is F, these cumulative probabilities are denoted by

$$P_{ij(c)} = F(\gamma_c - \theta_{ij}), \quad c = 1, \cdots, C-1,$$

where $\gamma_0 = -\infty$ and $\gamma_C = +\infty$. Equivalently, we can write the model as a cumulative model

$$G\left[P_{ij(c)}\right] = \gamma_c - \theta_{ij},$$

where $G = F^{-1}$ is the link function.

If ε_{ij} follows the logistic distribution, this results in the multilevel ordered logistic regression model, also called the multilevel ordered logit model or multilevel proportional odds model. If ε_{ij} has the standard normal distribution, this leads to the multilevel ordered probit model. The differences between these two models are minor and the choice between them is a matter of fit and convenience.

Assuming the distribution of the error term ε_{ij} of the latent response y^*_{ij} to be logistic, the cumulative probability function of y_{ij} will be written as

$$\begin{aligned} P_{ij(c)} &= Pr(\varepsilon_{ij} \leq \gamma_c - \theta_{ij}) \\ &= \frac{exp\left(\gamma_c - \theta_{ij}\right)}{1 + exp\left(\gamma_c - \theta_{ij}\right)}. \end{aligned}$$

The idea of cumulative probabilities leads naturally to the cumulative logit model

$$\begin{aligned} log\left[\frac{P_{ij(c)}}{1 - P_{ij(c)}}\right] &= log\left[\frac{Pr\left(y_{ij} \leq c\right)}{Pr\left(y_{ij} > c\right)}\right] \\ &= \gamma_c - \theta_{ij}, \qquad c = 1, \cdots, C-1, \end{aligned}$$

with $(C-1)$ strictly increasing model thresholds γ_c (i.e., $\gamma_1 < \gamma_2 ... < \gamma_{C-1}$). In this case, the observed ordinal outcome $y_{ij} = c$ if $\gamma_{c-1} \leq y^*_{ij} < \gamma_c$ for the latent variable (with $\gamma_0 = -\infty$ and $\gamma_C = +\infty$). As in the binary case, it is common to set one threshold to zero to fix the location of the latent variable. Typically, this is done in terms of the first threshold (i.e., $\gamma_1 = 0$).

4.3 Level-1 Model

With explanatory variables and random intercepts the level-1 model becomes

$$log\left[\frac{Pr\left(y_{ij} \leq c \mid \mathbf{x}_{ij}, \beta_{0j}\right)}{1 - Pr\left(y_{ij} \leq c \mid \mathbf{x}_{ij}, \beta_{0j}\right)}\right] = \gamma_c - \left(\beta_{0j} + \sum_{p=1}^{P} \beta_p x_{pij}\right),$$

where γ_c is the threshold parameter for category $c = 1, \cdots, C-1$.

Since the regression coefficients β do not carry the c subscript, they do not vary across categories. Thus, the relationship between the explanatory variables and the cumulative logits does not depend on c. This assumption of identical odds ratios across the $(C-1)$ partitions of the original ordinal outcome is called the proportional odds assumption (McCullagh, 1980). As written above, a positive coefficient for a regressor indicates that, as values of the regressor increase, so do the odds that the response is greater than or equal to c, for any $c = 1, \cdots, C-1$.

Although this is a natural way of writing the model, because it means that, for a positive β, as x increases so does the value of y^*, it is not the only way of writing the model. In particular, the model is sometimes written as

$$log\left[\frac{Pr\left(y_{ij} \leq c \mid \mathbf{x}_{ij}, \beta_{0j}\right)}{1 - Pr\left(y_{ij} \leq c \mid \mathbf{x}_{ij}, \beta_{0j}\right)}\right] = \gamma_c + \left(\beta_{0j} + \sum_{p=1}^{P} \beta_p x_{pij}\right),$$

in which case the regression parameters β are identical in magnitude but of opposite sign (see, eg. Raudenbush and Bryk, 2002).

4.4 Level-2 Model

The level-2 model has the usual form

$$\beta_{0j} = \gamma_{00} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + u_{0j},$$

where the random effects u_{0j} are normally distributed.

Note that the model which includes the intercept parameter γ_{00} and the threshold γ_1 is not identifiable. Let us consider a simple intercept model with no explanatory variables. For the first category we have

$$log\left[\frac{Pr\left(y_{ij} \leq 1 \mid u_{0j}\right)}{1 - Pr\left(y_{ij} \leq 1 \mid u_{0j}\right)}\right] = \gamma_1 - \left(\gamma_{00} + u_{0j}\right).$$

From this equation, it is apparent that parameters γ_1 and γ_{00} cannot be estimated separately and therefore those parameters are not identifiable. For identification, the first threshold γ_1 or the intercept γ_{00} may be fixed at zero. The Sabre syntax uses $\gamma_{00} = 0$.

4.5 Dichotomization of Ordered Categories

Models for ordered categorical outcomes are more complicated to fit and to interpret than models for dichotomous outcomes. Therefore it can make sense also to analyze the data after dichotomizing the outcome variable whilst retaining the ordinality of the response categories. For example, if there are 3 outcomes, one could analyze the dichotomization {1} versus {2, 3} and also {1, 2} versus {3}. Each of these analyses separately is based, of course, on less information, but may be easier to carry out and to interpret than an analysis of the original ordinal outcome.

4.6 Likelihood

$$L\left(\gamma, \sigma_{\varepsilon}^{2}, \sigma_{u_0}^{2} | \mathbf{y}, \mathbf{x}, \mathbf{z}\right) = \prod_{j} \int_{-\infty}^{+\infty} \prod_{i} g\left(y_{ij} | \mathbf{x}_{ij}, \mathbf{z}_{j}, u_{0j}\right) f\left(u_{0j}\right) du_{0j},$$

where

$$\begin{aligned} g\left(y_{ij} | \mathbf{x}_{ij}, \mathbf{z}_{j}, u_{0j}\right) &= \prod_{c} Pr(y_{ij} = c)^{y_{ijc}}, \\ &= \prod_{c} \left(P_{ij(c)} - P_{ij(c-1)}\right)^{y_{ijc}}, \end{aligned}$$

and $y_{ijc} = 1$, if $y_{ij} = c$, 0 otherwise,

$$\begin{aligned} P_{ij(c)} &= Pr\left(\varepsilon_{ij} \leq \left(\gamma_c - \left\{\gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + u_{0j}\right\}\right)\right) \\ &= F\left(\gamma_c - \left\{\gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + u_{0j}\right\}\right), \end{aligned}$$

where $F(\cdot)$ is the cumulative distribution function of ε_{ij} and

$$f\left(u_{0j}\right) = \frac{1}{\sqrt{2\pi}\sigma_{u_0}} \exp\left(-\frac{u_{0j}^{2}}{2\sigma_{u_0}^{2}}\right).$$

Sabre evaluates the integral $L\left(\gamma, \sigma_{\varepsilon}^{2}, \sigma_{u_0}^{2} | \mathbf{y}, \mathbf{x}, \mathbf{z}\right)$ for the ordered response model using normal Gaussian quadrature or adaptive Gaussian quadrature (numerical integration). There is not an analytic solution for this integral with normally distributed u_{0j}.

A cross-sectional example on teachers in schools (level 2) will be demonstrated.

4.7 Example C4. Ordered Response Model of Teacher's Commitment to Teaching

Rowan, Raudenbush and Cheong (1993) analysed data from a 1990 survey of teachers working in 16 public schools in California and Michigan. The schools were specifically selected to vary in terms of size, organizational structure, and urban versus suburban location. The survey asked the following question: if you could go back to college and start all over again, would you again choose teaching as a profession?' Possible responses were: 1 = *yes*; 2 = *not sure*; 3 = *no*. We take the teachers' response to this question as the response variable and try to establish if characteristics of the teachers and school help to predict their response to this question. We estimate 2 models, the first (on `teacher1.dta,` `#` observations = 661, `#` cases = 16) without covariates the second with. Because there are missing values in the covariates, the second data set (`teacher2.dta,#` observations = 650, `#` cases = 16) has fewer observations.

4.7.1 Reference

Rowan, B., Raudenbush, S., and Cheong, Y. (1993). Teaching as a non-routine task: implications for the organizational design of schools, Educational Administration Quarterly, 29(4), 479-500.

4.7.2 Data description for `teacher1.dta` and `teacher2.dta`

Number of observations in `teacher1.dta` (rows): 661
Number of observations in `teacher2.dta` (rows): 650
Number of level-2 cases: 16

4.7.3 Variables

We use a subset of the data with the following variables:
`tcommit`: the three-category measure of teacher commitment
`taskvar`: teachers' perception of task variety, which assesses the extent to which teachers followed the same teaching routines each day, performed the same tasks each day, had something new happening in their job each day, and liked the variety present in their work
`tcontrol`: a school-level variable, which is a measure of teacher control. This variable was constructed by aggregating nine item scale scores of teachers within a school. It indicates teacher control over school policy issues such as student behaviour codes, content of in-service programmes, student grouping, school curriculum, and text selection; and control over classroom issues such as teaching content and techniques, and amount of homework assigned.
`schlid`: school identifier

tcommit	taskvar	tcontrol	schlid
1	-0.26	-0.02	1
1	0.57	-0.02	1
1	0.13	-0.02	1
2	-0.26	-0.02	1
3	-1.10	-0.02	1
1	0.53	-0.02	1
2	0.61	-0.02	1
1	0.57	-0.02	1
1	-0.26	-0.02	1
3	-0.22	-0.02	1
3	-2.77	-0.02	1
2	0.57	-0.02	1
1	0.97	-0.02	1
1	1.01	-0.02	1
3	0.57	-0.02	1
1	-0.18	-0.02	1
2	-0.30	-0.02	1
1	-0.26	-0.02	1
3	-0.58	-0.02	1
1	-1.93	-0.02	1
1	0.17	-0.02	1

First few lines of `teacher2.dta`

The response variable `tcommit` takes on the value of $k = 1, 2, 3$. In the absence of explanatory variables and random intercepts, these values occur with probabilities

$$\begin{aligned} p_{ij(1)} &= Pr(y_{ij} = 1) = Pr(\text{"Yes"}), \\ p_{ij(2)} &= Pr(y_{ij} = 2) = Pr(\text{"Not sure"}), \\ p_{ij(3)} &= Pr(y_{ij} = 3) = Pr(\text{"No"}). \end{aligned}$$

To assess the magnitude of variation among schools in the absence of explanatory variables, we specify a simple 1-level model. This model has only the thresholds and the school-specific intercepts as fixed effects:

$$log\left[\frac{Pr(y_{ij} \leq c \mid \beta_{0j})}{Pr(y_{ij} > c \mid \beta_{0j})}\right] = \gamma_c - \beta_{0j}, \qquad c = 1, 2.$$

The 2-level model is

$$\beta_{0j} = \gamma_{00} + u_{0j},$$

though the model is identifiable as long as the parameter γ_{00} is set to zero. This reduces the 2-level model to $\beta_{0j} = u_{0j}$. Rather than treat the school-specific intercepts β_{0j} as fixed effects, we now regard the school-specific intercepts u_{0j} as random effects with variance $\sigma^2_{u_0}$. Next, we consider the introduction of explanatory variables into this model. Rowan, Raudenbush, and Cheong (1993) hypothesized that teachers would express high levels of commitment if they had a job with a high degree of task variety and also experienced a high degree of control over school policies and teaching conditions. Conceptually, task variety varies at the teacher level, while teacher control varies at the school level.

The level-1 model is

$$log\left[\frac{Pr(y_{ij} \leq c \mid \mathbf{x}_{ij}, \beta_{0j})}{Pr(y_{ij} > c \mid \mathbf{x}_{ij}, \beta_{0j})}\right] = \gamma_c - (\beta_{0j} + \beta_{1j}\texttt{taskvar}_{ij}),$$

while the level-2 model is

$$\beta_{0j} = \gamma_{01}(\texttt{tcontrol})_j + u_{0j},$$
$$\beta_{1j} = \gamma_{10}.$$

The combined model is

$$log\left[\frac{Pr(y_{ij} \leq c \mid \mathbf{x}_{ij}, \mathbf{z}_j, u_{0j})}{Pr(y_{ij} > c \mid \mathbf{x}_{ij}, \mathbf{z}_j, u_{0j})}\right] = \gamma_c - (\gamma_{01}\texttt{tcontrol}_j + \gamma_{10}\texttt{taskvar}_{ij} + u_{0j}).$$

To fit these models with and without explanatory variables we use Sabre 5.0.

4.7.4 Sabre commands

```
log using teacher_s.log, replace
set more off
use teacher1
sabre, data tcommit tcontrol schlid
sabre tcommit tcontrol schlid, read
sabre, case schlid
sabre, yvar tcommit
sabre, ordered y
sabre, constant cons
sabre, lfit
sabre, dis m
sabre, dis e
sabre, fit
sabre, dis m
sabre, dis e
clear
use teacher2
sabre, data tcommit taskvar tcontrol schlid
sabre tcommit taskvar tcontrol schlid, read
sabre, case schlid
sabre, yvar tcommit
sabre, ordered y
sabre, constant cons
sabre, lfit tcontrol taskvar
sabre, dis m
sabre, dis e
sabre, fit tcontrol taskvar
sabre, dis m
sabre, dis e
log close
```

```
clear
exit
```

4.7.5 Sabre log file

```
(Random Effects Model)


    Univariate model
    Standard ordered logit
    Gaussian random effects

    Number of observations             =     661
    Number of cases                    =      16

    X-var df           =     0
    Cutpoint df        =     2
    Scale df           =     1

    Log likelihood =     -662.66290     on     658 residual degrees of freedom


    Parameter              Estimate         Std. Err.
    ___________________________________________________
    (intercept)            -1.2480          0.13296
    cut1                   -1.0309          0.78614E-01
    cut2                    0.0000           0.0000
    scale                  0.33527          0.13507


(Random Effects Model)


    Univariate model
    Standard ordered logit
    Gaussian random effects

    Number of observations             =     650
    Number of cases                    =      16

    X-var df           =     2
    Cutpoint df        =     2
    Scale df           =     1

    Log likelihood =     -634.05978     on     645 residual degrees of freedom


    Parameter              Estimate         Std. Err.
    ___________________________________________________
    (intercept)            -1.2499          0.95459E-01
    tcontrol               -1.5410          0.36060
    taskvar               -0.34881          0.87745E-01
    cut1                   -1.0570          0.80838E-01
    cut2                    0.0000           0.0000
```

```
scale                    0.16144E-05      0.17659
```

4.7.6 Discussion

For the model without covariates, the results indicate that the estimated values of the threshold parameters are 0.217 (γ_1), 1.248 (γ_2), and that the estimate of the variance of the school-specific intercepts, $\sigma^2_{u_0}$, is $(0.33527)^2 = 0.11241$.

The model formulation summarizes the two equations as

$$log\left[\frac{Pr(y_{ij} \leq 1 \mid u_{0j})}{Pr(y_{ij} > 1 \mid u_{0j})}\right] = 0.217 - u_{0j},$$
$$log\left[\frac{Pr(y_{ij} \leq 2 \mid u_{0j})}{Pr(y_{ij} > 2 \mid u_{0j})}\right] = 1.248 - u_{0j}.$$

For the model with explanatory variables included, the two equations summarizing these results are:

$$log\left[\frac{Pr(y_{ij} \leq 1 \mid \mathbf{x}_{ij}, \mathbf{z}_j, u_{0j})}{Pr(y_{ij} > 1 \mid \mathbf{x}_{ij}, \mathbf{z}_j, u_{0j})}\right] = 0.193 - [(-0.349\texttt{taskvar}_{ij} - 1.541\texttt{tcontrol}_j + u_{0j})]$$
$$= 0.193 + 0.349\texttt{taskvar}_{ij} + 1.541\texttt{tcontrol}_j - u_{0j},$$

$$log\left[\frac{Pr(y_{ij} \leq 2 \mid \mathbf{x}_{ij}, \mathbf{z}_j, u_{0j})}{Pr(y_{ij} > 2 \mid \mathbf{x}_{ij}, \mathbf{z}_j, u_{0j})}\right] = 1.248 + 0.349\texttt{taskvar}_{ij} + 1.541\texttt{tcontrol}_j - u_{0j}.$$

The results indicate that, within schools, `taskvar` is significantly related to commitment ($\gamma_{10} = 0.349$, $ztest = 3.98$); between schools, `tcontrol` is also strongly related to commitment ($\gamma_{01} = 1.541, ztest = 4.27$). Inclusion of `tcontrol` reduced the point estimate of the between-school variance to 0.000. This suggests that we do not need random effects in the model with explanatory variables. The model without the random effect u_{0j} will be

$$log\left[\frac{Pr(y_{ij} \leq c \mid \mathbf{x}_{ij}, \mathbf{z}_j)}{Pr(y_{ij} > c \mid \mathbf{x}_{ij}, \mathbf{z}_j)}\right] = \gamma_c + 0.349\texttt{taskvar}_{ij} + 1.541\texttt{tcontrol}_j, c = 1, 2.$$

For further discussion on ordered response models with random intercepts see: Rabe-Hesketh and Skrondal (2005), and Wooldridge (2002).

4.8 Exercises

There are three ordered response model exercises, namely C4, L5 and L6.

4.9 References

McCullagh, P., (1980) 'Regression Models for Ordinal Data (with discussion)', Journal of the Royal Statistical Society B, vol. 42, 109 - 142.

Rabe-Hesketh, S., and Skrondal, A., (2005), Multilevel and Longitudinal Modelling using Stata, Stata Press, Stata Corp, College Station, Texas.

Wooldridge, J., M., (2006), Introductory Econometrics: A Modern Approach. Third edition. Thompson, Australia.

Chapter 5

Multilevel Poisson Models

5.1 Introduction

Another important type of discrete data is count data. For example, for a population of road crossings one might count the number of accidents in one year; or for a population of doctors, one could count how often in one year they are confronted with a certain medical problem. The set of possible outcomes of count data is the set of natural numbers: $0, 1, 2, \cdots$. The standard distribution for counts is the Poisson distribution. Suppose y_{ij} to be a variable distributed randomly as $Poisson(\mu_{ij})$. Then we write

$$Pr(y_{ij}) = \frac{exp(-\mu_{ij})\mu_{ij}^{y_{ij}}}{y_{ij}!}, \quad y_{ij} = 0, 1, \cdots .$$

The Poisson distribution has some properties that we can make use of when modelling our data. For example, the expected or mean value of y is equal to the variance of y, so that

$$E(y_{ij}) = var(y_{ij}) = \mu_{ij}.$$

When we have Poisson distributed data, it is usual to use a logarithmic transformation to model the mean, i.e. $log(\mu_{ij})$. This is the natural parameter for modelling the Poisson distribution. There is no theoretical restriction, however, on using other transformations of μ_{ij}, so long as the mean is positive, as discussed in Dobson (1991).

Further, if the counts tend to be large, their distribution can be approximated by a continuous distribution. If all counts are large enough, then it is advisable to use the square root of the counts as the response variable and then fit the model. The reason why this is a good approach resides in the fact that the square root transformation succeeds very well in transforming the Poisson distribution to an approximately homoscedastic normal distribution (the square root is the so-called variance-stabilizing transformation for the Poisson distribution).

If all or some of the counts are small, a normal distribution will not be satisfactory.

5.2 Poisson Regression Models

In Poisson regression it is assumed that the response variable y_{ij} has a Poisson distribution given the explanatory variables $x_{1ij}, x_{2ij}, \cdots, x_{pij}$,

$$y_{ij}|x_{1ij}, x_{2ij}, \cdots, x_{pij} \sim Poisson(\mu_{ij}),$$

where the log of the mean μ_{ij} is assumed to be a linear function of the explanatory variables. That is,

$$log(\mu_{ij}) = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \cdots + \beta_{pj}x_{pij},$$

which implies that μ_{ij} is the exponential function of independent variables,

$$\mu_{ij} = exp\left(\beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \cdots + \beta_{pj}x_{pij}\right).$$

In models for counts it is quite usual that there is a variable M_{ij} that is known to be proportional to the expected counts. For example, if the count y_{ij} is the number of events in some time interval of non-constant length m_{ij}, it is often natural to assume that the expected count is proportional to this length of the time period. In order to let the expected count be proportional to M_{ij}, there should be a term $log(m_{ij})$ in the linear model for $log(\mu_{ij})$, with a regression coefficient fixed to 1. Such a term is called an *offset* in the linear model (see e.g., McCullagh and Nelder, 1989; Goldstein, 2003). Therefore, the Poisson regression model can be written in the following form:

$$log(\mu_{ij}) = log\left(m_{ij}\right) + \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \cdots + \beta_{pj}x_{pij}.$$

The $log(\mu_{ij}/m_{ij})$ is modelled now as a linear function of explanatory variables.

5.3 The Two-Level Poisson Model

Let y_{ij} be the count for level-1 unit i in level-2 unit j, and μ_{ij} be the expected count, given that level-1 unit i is in level-2 unit j and given the values of the explanatory variables. Then μ_{ij} is necessarily a non-negative number, which could lead to difficulties if we considered linear models for this value. The natural logarithm is mostly used as the link function for expected counts. For single-level data this leads to the Poisson regression model which is a linear model for the natural logarithm of the counts, $log(\mu_{ij})$. For multilevel data, hierarchical linear models are considered for the logarithm of μ_{ij}.

5.4 Level-1 Model

Consider a two-level multilevel Poisson model by assuming the level-1 units i are nested within level-2 units j. Using the logarithmic transformation, the level-1 model with P explanatory variables $x_1, \cdots, x_P$ may be written as

$$y_{ij} \sim Poisson(\mu_{ij}),$$

$$log(\mu_{ij}) = log(m_{ij}) + \beta_{0j} + \sum_{p=1}^{P} \beta_{pj} x_{pij},$$

where β_{0j} is an intercept parameter, and β_{pj}, $p = 1, \cdots, P$, are slope parameters associated with explanatory variables x_{pij}. The term $log(m_{ij})$ is included in the model as an offset.

5.5 Level-2 Model: The Random Intercept Model

The level-2 model has the same form as the level-2 model in the linear model, binary and ordinal response models. Consider for example the random intercept model formulated as a regression model plus a random intercept for the logarithm of the expected count. As we are limited to random intercepts we have:

$$\beta_{pj} = \gamma_{p0},$$

$$\beta_{0j} = \gamma_{00} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + u_{0j},$$

so that

$$log(\mu_{ij}) = log(m_{ij}) + \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + u_{0j}.$$

The variance of the random intercept is denoted again by $\sigma^2_{u_0}$.

To transform the linear model back to the expected counts, the inverse transformation of the natural logarithm must be used. Therefore, the explanatory variables and the level-two random effects in the (additive) multilevel Poisson regression model have multiplicative effects on the expected counts.

5.6 Likelihood

$$L\left(\gamma, \sigma_{u_0}^2 | \mathbf{y}, \mathbf{x}, \mathbf{z}\right) = \prod_j \int_{-\infty}^{+\infty} \prod_i g\left(y_{ij} | \mathbf{x}_{ij}, \mathbf{z}_{j,} u_{0j}\right) f\left(u_{0j}\right) du_{0j},$$

where

$$g\left(y_{ij} | \mathbf{x}_{ij}, \mathbf{z}_{j,} u_{0j}\right) = \frac{exp(-\mu_{ij})\mu_{ij}^{y_{ij}}}{y_{ij}!},$$

and

$$f\left(u_{0j}\right) = \frac{1}{\sqrt{2\pi}\sigma_{u_0}} \exp\left(-\frac{u_{0j}^2}{2\sigma_{u_0}^2}\right).$$

Sabre evaluates the integral $L\left(\gamma, \sigma_{u_0}^2 | \mathbf{y}, \mathbf{x}, \mathbf{z}\right)$ for the Poisson model using normal Gaussian quadrature or adaptive Gaussian quadrature (numerical integration). There is not an analytic solution for this integral with normally distributed u_{0j}.

5.7 Example C5. Poisson Model of Prescribed Medications

Cameron and Trivedi (1988) use various forms of overdispersed Poisson model to study the relationship between type of health insurance and various responses which measure the demand for health care, such as the total number of prescribed medications used in the past 2 days. The data set they use in this analysis is from the Australian Health survey for 1977-1978. A copy of the original data set and further details about the variables in `racd.dta` can be obtained from
`http://cameron.econ.ucdavis.edu/racd/racddata.html`.

5.7.1 References

Cameron, A.C., Trivedi, P.K., Milne, F., Piggott, J., (1988) A microeconometric model of the demand for Health Care and Health Insurance in Australia, Review of Economic Studies, 55, 85-106.

Cameron, A.C., Trivedi, P.K (1998), Regression Analysis of Count Data, Econometric Society Monograph No.30, Cambridge University Press

5.7.2 Data description for `racd.dta`

Number of observations (rows): 5190
Number of level-2 cases: 5190

5.7.3 Variables

`sex`: 1 if respondent is female, 0 if male
`age`: respondent's age in years divided by 100
`agesq`: age squared
`income`: respondent's annual income in Australian dollars divided by 1000
`levyplus`: 1 if respondent is covered by private health insurance fund for private patients in public hospital (with doctor of choice), 0 otherwise
`freepoor`: 1 if respondent is covered by government because low income, recent immigrant, unemployed, 0 otherwise
`freerepa`: 1 if respondent is covered free by government because of old-age or disability pension, or because invalid veteran or family of deceased veteran, 0 otherwise
`illness`: number of illnesses in past 2 weeks, with 5 or more weeks coded as 5
`actdays`: number of days of reduced activity in past two weeks due to illness or injury
`hscore`: respondent's general health questionnaire score using Goldberg's method, high score indicates poor health

`chcond1`: 1 if respondent has chronic condition(s) but is not limited in activity, 0 otherwise
`chcond2`: 1 if respondent has chronic condition(s) and is limited in activity, 0 otherwise
`dvisits`: number of consultations with a doctor or specialist in the past 2 weeks
`nondocco`: number of consultations with non-doctor health professionals (chemist, optician, physiotherapist, social worker, district community nurse, chiropodist or chiropractor) in the past 2 weeks
`hospadmi`: number of admissions to a hospital, psychiatric hospital, nursing or convalescent home in the past 12 months (5 or more admissions coded as 5)
`hospdays`: number of nights in a hospital, etc. during most recent admission, in past 12 months
`medicine`: total number of prescribed and nonprescribed medications used in past 2 days
`prescrib`: total number of prescribed medications used in past 2 days
`nonpresc`: total number of nonprescribed medications used in past 2 days
`constant`: 1 for all observations
`id`: ij

Like Cameron and Trivedi we take `prescrib` to be the Poisson response variable and model it with a random intercept and a range of explanatory variables.

sex	age	agesq	income	levyplus	freepoor	freerepa	illness	actdays	hscore	chcond1	chcond2	dvisits	nondocco	hospadmi	hospdays	medicine	prescrib	nonpresc	constant	id
1	0.19	0.04	0.55	1	0	0	1	4	1	0	0	1	0	0	0	1	1	0	1	1
1	0.19	0.04	0.45	1	0	0	1	2	1	0	0	1	0	0	0	2	1	1	1	2
0	0.19	0.04	0.90	0	0	0	3	0	0	0	0	1	0	1	4	2	1	1	1	3
0	0.19	0.04	0.15	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	4
0	0.19	0.04	0.45	0	0	0	2	5	1	1	0	1	0	0	0	3	1	2	1	5
1	0.19	0.04	0.35	0	0	0	5	1	9	1	0	1	0	0	0	1	1	0	1	6
1	0.19	0.04	0.55	0	0	0	4	0	2	0	0	1	0	0	0	0	0	0	1	7
1	0.19	0.04	0.15	0	0	0	3	0	6	0	0	1	0	0	0	1	1	0	1	8
1	0.19	0.04	0.65	1	0	0	2	0	5	0	0	1	0	0	0	1	0	1	1	9
0	0.19	0.04	0.15	1	0	0	1	0	0	0	0	1	0	0	0	1	1	0	1	10
0	0.19	0.04	0.45	0	0	0	1	0	0	0	0	1	0	0	0	1	1	0	1	11
0	0.19	0.04	0.25	0	0	1	2	0	2	0	0	1	0	1	80	1	1	0	1	12
0	0.19	0.04	0.55	0	0	0	3	13	1	1	0	2	0	0	0	0	0	0	1	13
0	0.19	0.04	0.45	0	0	0	4	7	6	1	0	1	0	0	0	0	0	0	1	14
0	0.19	0.04	0.25	1	0	0	3	1	0	1	0	1	0	0	0	2	2	0	1	15
0	0.19	0.04	0.55	0	0	0	2	0	7	0	0	1	0	0	0	3	2	1	1	16
0	0.19	0.04	0.45	1	0	0	1	0	5	0	0	2	0	0	0	1	1	0	1	17
1	0.19	0.04	0.45	0	0	0	1	1	0	1	0	1	0	0	0	1	1	0	1	18
1	0.19	0.04	0.45	1	0	0	1	0	0	0	0	2	0	0	0	1	1	0	1	19
1	0.19	0.04	0.35	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	20
1	0.19	0.04	0.45	1	0	0	1	3	0	0	0	1	0	0	0	0	0	0	1	21
1	0.19	0.04	0.35	1	0	0	1	0	1	0	0	1	0	0	0	2	1	1	1	22
0	0.19	0.04	0.45	1	0	0	2	2	0	0	0	1	0	0	0	0	0	0	1	23
0	0.19	0.04	0.55	0	0	0	2	14	2	0	0	1	0	0	0	1	1	0	1	24
1	0.19	0.04	0.25	0	0	1	2	14	11	0	1	1	0	1	11	5	5	0	1	25
1	0.19	0.04	0.15	0	1	0	1	2	6	1	0	1	0	0	0	2	2	0	1	26
1	0.19	0.04	0.55	0	0	0	2	5	6	0	0	1	0	0	0	1	1	0	1	27

First few lines and columns of `racd.dta`

5.7.4 Sabre commands

```
log using prescribe_s.log, replace
set more off
use racd
#delimit ;
sabre, data sex age agesq income levyplus freepoor freerepa illness actdays
       hscore chcond1 chcond2 dvisits nondocco hospadmi hospdays medicine
       prescrib nonpresc constant id;
```

```
sabre sex age agesq income levyplus freepoor freerepa illness actdays hscore
      chcond1 chcond2 dvisits nondocco hospadmi hospdays medicine prescrib
      nonpresc constant id, read;
#delimit cr
sabre, case id
sabre, yvar prescrib
sabre, family p
sabre, constant cons
#delimit ;
sabre, lfit sex age agesq income levyplus freepoor freerepa illness actdays
            hscore chcond1 chcond2 cons;
#delimit cr
sabre, dis m
sabre, dis e
#delimit ;
sabre, fit sex age agesq income levyplus freepoor freerepa illness actdays
           hscore chcond1 chcond2 cons;
#delimit cr
sabre, dis m
sabre, dis e
log close
clear
exit
```

5.7.5 Sabre log file

```
(Standard Homogenous Model)

    Parameter             Estimate         Std. Err.
    ___________________________________________________
    (intercept)           -2.7412          0.12921
    sex                    0.48377         0.36639E-01
    age                     2.6497         0.61491
    agesq                 -0.88778         0.64292
    income                -0.44661E-02     0.55766E-01
    levyplus               0.28274         0.52278E-01
    freepoor              -0.45680E-01     0.12414
    freerepa               0.29584         0.59667E-01
    illness                0.20112         0.10530E-01
    actdays                0.29261E-01     0.36746E-02
    hscore                 0.20103E-01     0.63664E-02
    chcond1                0.77565         0.46130E-01
    chcond2                 1.0107         0.53895E-01


(Random Effects Model)

    Parameter             Estimate         Std. Err.
    ___________________________________________________
    (intercept)           -2.8668          0.14908
    sex                    0.56080         0.43164E-01
    age                     2.0861         0.73513
    agesq                 -0.26325         0.78264
```

```
income               0.30450E-01      0.65221E-01
levyplus             0.27060          0.58009E-01
freepoor            -0.61759E-01      0.13676
freerepa             0.29172          0.69172E-01
illness              0.20914          0.13260E-01
actdays              0.34688E-01      0.49475E-02
hscore               0.21604E-01      0.81424E-02
chcond1              0.77394          0.50771E-01
chcond2               1.0245          0.62314E-01
scale                0.52753          0.27207E-01


Univariate model
Standard Poisson
Gaussian random effects

Number of observations            =    5190
Number of cases                   =    5190

X-var df          =    13
Scale df          =     1

Log likelihood =     -5443.3311     on    5176 residual degrees of freedom
```

5.7.6 Discussion

This shows that even with a range of explanatory variables included in the model, there is still a highly significant amount of between-respondent variation in the total number of prescribed medications used in the past 2 days, as indicated by the scale parameter estimate of 0.52753 (s.e.0.027207).

The random effect model parameter estimates differ slightly from those of the homogeneous model. If the random effect model is the true model, then asymptotically both the homogeneous and random effect model estimates will tend to the same limit. As expected the standard errors for the random effect model estimates are larger than those of the homogeneous model.

In this analysis we only have 1 response for each subject. We do not need multiple responses to identify the extra variation in Poisson counts. However, having multiple responses for each subject would give two ways to identify the extra variation: (1) from the extra variation in each of a subject's responses and (2) from the correlation between the different responses of each subject.

For further discussion on Poisson models with random intercepts see: Cameron and Trivedi (1998), Rabe-Hesketh and Skrondal (2005) and Wooldridge (2002).

5.8 Exercises

There are two Poisson response model exercises, namely C5 and L8.

5.9 References

Cameron, A., Trivedi, P.K., (1998), Regression Analsyis of Count Data, Cambridge, Cambridge University Press.

Rabe-Hesketh, S., and Skrondal, A., (2005), Multilevel and Longitudinal Modelling using Stata, Stata Press, Stata Corp, College Station, Texas.

Wooldridge, J. M. (2002), Econometric Analysis of Cross Section and Panel Data, MIT Press, Cambridge Mass.

Chapter 6

Two-Level Generalised Linear Mixed Models

6.1 Introduction

The main models we have considered so far, namely linear, binary response and Poisson models, are special cases of the generalised linear model (GLM) or exponential family. It will help us in considering extensions of these models to 3 levels, and to multivariate responses, if we can start to write each of the models using GLM notation. In generalised linear models, the explanatory variables and the random effects (for a 2-level model these are x_{ij},z_j and u_{0j}) affect the response (for a 2-level model this is y_{ij}) via the linear predictor (θ_{ij}), where

$$\theta_{ij} = \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + u_{0j}.$$

The GLM is obtained by specifying some function of the response (y_{ij}) conditional on the linear predictor and other parameters, i.e.

$$g\left(y_{ij} \mid \theta_{ij}, \phi\right) = \exp\left\{\left[y_{ij}\theta_{ij} - b\left(\theta_{ij}\right)\right]/\phi + c\left(y_{ij}, \phi\right)\right\},$$

where ϕ is the scale parameter, $b\left(\theta_{ij}\right)$ is a function that gives the conditional mean (μ_{ij}) and variance of y_{ij}, namely

$$E\left[y_{ij} \mid \theta_{ij}, \phi\right] = \mu_{ij} = b^{'}\left(\theta_{ij}\right),$$
$$Var\left[y_{ij} \mid \theta_{ij}, \phi\right] = \phi b^{''}\left(\theta_{ij}\right).$$

In GLMs the mean and variance are related so that

$$Var\left[y_{ij} \mid \theta_{ij}, \phi\right] = \phi b^{''}\left(b^{'-1}\left(\theta_{ij}\right)\right) = \phi V\left[\mu_{ij}\right].$$

$V(\mu_{ij})$ is called the variance function. The function $b^{'-1}(\theta_{ij})$ which expresses θ_{ij} as a function of μ_{ij} is called the link function, and $b^{'}(\theta_{ij})$ is the inverse link function.

Both $b(\theta_{ij})$ and $c(y_{ij}, \phi)$ differ for different members of the exponential family.

6.2 The Linear Model

If we rewrite the linear model from an earlier section as

$$\begin{aligned} g(y_{ij}|\mathbf{x}_{ij}, \mathbf{z}_j, u_{0j}) &= g(y_{ij} \mid \theta_{ij}, \phi) \\ &= \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} \exp\left(-\frac{[y_{ij} - \mu_{ij}]^2}{2\sigma_\varepsilon^2}\right), \end{aligned}$$

then we can write

$$g(y_{ij} \mid \theta_{ij}, \phi) = \exp\left\{\frac{1}{2\sigma_\varepsilon^2}\left(y_{ij}\mu_{ij} - \frac{\mu_{ij}^2}{2}\right) + \left(\frac{\ln(2\pi\sigma_\varepsilon)}{2} - \frac{y_{ij}^2}{2\sigma_\varepsilon^2}\right)\right\},$$

so that

$$\begin{aligned} \theta_{ij} &= \mu_{ij}, \\ \phi &= \sigma_\varepsilon^2, \\ b(\theta_{ij}) &= \frac{\theta_{ij}^2}{2}, \\ c(y_{ij}, \phi) &= \frac{\ln(2\pi\sigma_\varepsilon)}{2} - \frac{y_{ij}^2}{2\sigma_\varepsilon^2}. \end{aligned}$$

The mean (μ_{ij}) and variance functions are

$$\begin{aligned} \mu_{ij} &= \theta_{ij}, \\ V[\mu_{ij}] &= 1. \end{aligned}$$

Note that in the linear model, the mean and variance are not related as

$$\phi V[\mu_{ij}] = \sigma_\varepsilon^2.$$

Also the link function is the identity as $\theta_{ij} = \mu_{ij}$. We define this model by Gaussian error g, identity link i.

6.3 Binary Response Models

If we rewrite the binary response model from an earlier section as

$$\begin{aligned} g\left(y_{ij}|\mathbf{x}_{ij},\mathbf{z}_j,u_{0j}\right) &= g\left(y_{ij}\mid\theta_{ij},\phi\right) \\ &= \mu_{ij}^{y_{ij}}\left(1-\mu_{ij}\right)^{1-y_{ij}}, \end{aligned}$$

then we can write

$$\begin{aligned} g\left(y_{ij}\mid\theta_{ij},\phi\right) &= \exp\left\{y_{ij}\ln\mu_{ij}+(1-y_{ij})\ln(1-\mu_{ij})\right\} \\ &= \exp\left\{y_{ij}\ln\left(\frac{\mu_{ij}}{(1-\mu_{ij})}\right)+\ln(1-\mu_{ij})\right\}, \end{aligned}$$

so that

$$\begin{aligned} \theta_{ij} &= \ln\left(\frac{\mu_{ij}}{(1-\mu_{ij})}\right), \\ \phi &= 1, \\ b\left(\theta_{ij}\right) &= \ln(1-\mu_{ij}), \\ c\left(y_{ij},\phi\right) &= 0. \end{aligned}$$

The mean (μ_{ij}) and variance functions are

$$\begin{aligned} \mu_{ij} &= \frac{\exp\left(\theta_{ij}\right)}{1+\exp\left(\theta_{ij}\right)}, \\ V\left[\mu_{ij}\right] &= \frac{\exp\left(\theta_{ij}\right)}{\left\{1+\exp\left(\theta_{ij}\right)\right\}^2}. \end{aligned}$$

Note that in the binary response model, the mean and variance are related as

$$\phi V\left[\mu_{ij}\right] = \mu_{ij}\left(1-\mu_{ij}\right).$$

Also $\theta_{ij} = \ln\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right)$, and the logit model (logit link) has

$$\mu_{ij} = \frac{\exp\left(\theta_{ij}\right)}{1+\exp\left(\theta_{ij}\right)}.$$

The probit model (probit link) has $\mu_{ij} = \Phi\left(\theta_{ij}\right)$, or $\Phi^{-1}\left(\mu_{ij}\right) = \theta_{ij}$, where $\Phi\left(.\right)$ is the standard normal cumulative distribution function. The complementary log log model (cloglog link) has $\theta_{ij} = \log\left\{-\log\left(1-\mu_{ij}\right)\right\}$, or $\mu_{ij} = 1-\exp\left(-\exp\theta_{ij}\right)$.

We define the binary response model with binomial error b, and logit, probit or cloglog link.

6.4 Poisson Model

If we rewrite the Poisson model from an earlier section as

$$\begin{aligned} g\left(y_{ij}|\mathbf{x}_{ij}, \mathbf{z}_j, u_{0j}\right) &= g\left(y_{ij} \mid \theta_{ij}, \phi\right) \\ &= \frac{exp(-\mu_{ij})\mu_{ij}^{y_{ij}}}{y_{ij}!}, \end{aligned}$$

then we can write

$$g\left(y_{ij} \mid \theta_{ij}, \phi\right) = \exp\left\{\left[y_{ij} \ln \mu_{ij} - \mu_{ij}\right] - \log y_{ij}!)\right\},$$

so that

$$\begin{aligned} \theta_{ij} &= \ln \mu_{ij}, \\ \phi &= 1, \\ b\left(\theta_{ij}\right) &= \mu_{ij} = \exp \theta_{ij}, \\ c\left(y_{ij}, \phi\right) &= -\log y_{ij}!. \end{aligned}$$

The mean (μ_{ij}) and variance functions are

$$\begin{aligned} \mu_{ij} &= \exp\left(\theta_{ij}\right), \\ V\left[\mu_{ij}\right] &= \mu_{ij}. \end{aligned}$$

Note that in the Poisson model, the mean and variance are related as

$$\phi V\left[\mu_{ij}\right] = \mu_{ij}.$$

The link function is the log link as $\theta_{ij} = \ln \mu_{ij}$. We define the Poisson model with Poisson error p, and logit g , probit p or cloglog c link.

6.5 Two-Level Generalised Linear Model Likelihood

We can now write the 2-level Generalised Linear Model (Generalised Linear Mixed Model) likelihood for the linear model, binary response and Poisson models in a general form, i.e.

$$L\left(\gamma, \phi, \sigma_{u_0}^2|\mathbf{y}, \mathbf{x}, \mathbf{z}\right) = \prod_j \int \prod_i g\left(y_{ij} \mid \theta_{ij}, \phi\right) f\left(u_{0j}\right) du_{0j},$$

where

$$g\left(y_{ij} \mid \theta_{ij}, \phi\right) = \exp\left\{\left[y_{ij}\theta_{ij} - b\left(\theta_{ij}\right)\right] / \phi + c\left(y_{ij}, \phi\right)\right\},$$

$$\theta_{ij} = \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + u_{0j},$$

and

$$f\left(u_{0j}\right) = \frac{1}{\sqrt{2\pi}\sigma_{u_0}} \exp\left(-\frac{u_{0j}^2}{2\sigma_{u_0}^2}\right).$$

For the linear model we have identity link , Gaussian (normal) error, for the binary model we have logit, probit, cloglog link, binomial error, and for the Poisson model we have log link and Poisson error. Sabre evaluates the integral $L\left(\gamma, \phi, \sigma_{u_0}^2 | \mathbf{y}, \mathbf{x}, \mathbf{z}\right)$ for the multilevel GLM model using normal Gaussian quadrature or adaptive Gaussian Quadrature (numerical integration).

For further discussion on GLMMs see Aitkin (1996, 1999)

6.6 References

Aitkin, M., (1996), A general maximum likelihood analysis of overdispersion in generalized linear models. Statistics and Computing, 6:251–262.

Aitkin, M., (1999), A general maximum likelihood analysis of variance components in generalized linear models. Biometrics, 55:218–234.

Chapter 7

Three-Level Generalised Linear Mixed Models

7.1 Introduction

The extension of the two-level regression model to three and more levels is reasonably straightforward. In this section we consider the three-level random intercept GLM.

7.2 Three-Level Random Intercept Models

In a previous example, data were used where students were nested within schools. The actual hierarchical structure of educational data is more usually students nested within classes nested within schools. For the time being we concentrate on 'simple' three-level hierarchical data structures. The response variable now needs to acknowledge the extra level and is denoted by y_{ijk}, referring to, e.g., the response of student i in class j in school k. More generally, one can talk about level-one unit i in level-two unit j in level-three unit k. The three-level model for such data with one level-1 explanatory variable may be formulated through the linear predictor. In this simple example we only use one level-1 covariate x_{ijk},so that

$$\theta_{ijk} = \beta_{0jk} + \beta_{1jk} x_{ijk},$$

where β_{0jk} is the intercept in level-two unit j within level-three unit k. For the intercept we have the level-two model,

$$\begin{aligned}\beta_{0jk} &= \delta_{00k} + u_{0jk}, \\ \beta_{1jk} &= \gamma_{100},\end{aligned}$$

where δ_{00k} is the average intercept in level-three unit k. For this average intercept we have the level-three model,

$$\delta_{00k} = \gamma_{000} + v_{00k},$$

and hence by substituting, the linear predictor takes the form

$$\theta_{ijk} = \gamma_{000} + \gamma_{100}x_{ijk} + v_{00k} + u_{0jk}.$$

7.3 Three-Level GLM

By using ijk subscripts for various terms of a GLM and by adding the level-3 explanatory covariates w_k and the level-2 explanatory variables z_{jk}, we get the 3-level GLM, where

$$g\left(y_{ijk} \mid \theta_{ijk}, \phi\right) = \exp\left\{\left[y_{ijk}\theta_{ijk} - b\left(\theta_{ijk}\right)\right]/\phi + c\left(y_{ijk}, \phi\right)\right\},$$

$$\theta_{ijk} = \gamma_{000} + \sum_{p=1}^{P}\gamma_{p00}x_{pijk} + \sum_{q=1}^{Q}\gamma_{0q0}z_{qjk} + \sum_{r=1}^{R}\gamma_{00r}w_{rk} + v_{00k} + u_{0jk}.$$

The conditional mean (μ_{ijk}) and variance of y_{ijk} become

$$E\left[y_{ijk} \mid \theta_{ijk}, \phi\right] = \mu_{ijk} = b^{'}\left(\theta_{ijk}\right),$$
$$Var\left[y_{ijk} \mid \theta_{ijk}, \phi\right] = \phi b^{''}\left(\theta_{ijk}\right),$$

and

$$Var\left[y_{ijk} \mid \theta_{ijk}, \phi\right] = \phi b^{''}\left(b^{'-1}\left(\theta_{ijk}\right)\right) = \phi V\left[\mu_{ijk}\right],$$

where $b\left(\theta_{ijk}\right)$ and $c\left(y_{ijk}, \phi\right)$ differ for different members of the exponential family.

For GLMs we can consider the covariances between the different linear predictors θ_{ijk} and $\theta_{i'jk}$ of different pupils i and i' in the same class of a given school and between different pupils j and j' in different classes of the same school, i.e. different linear predictors θ_{ijk} and $\theta_{i'j'k}$ of

$$covar\left(\theta_{ijk}, \theta_{i'jk} \mid x_{ijk}, z_{jk}, w_k\right) = \sigma_{u_0}^2 + \sigma_{v_{00}}^2,$$

$$covar\left(\theta_{ijk}, \theta_{i'j'k} \mid x_{ijk}, z_{jk}, w_k\right) = \sigma_{v_{00}}^2,$$

so that the covariance of different pupils in the same class in a given school is higher than that of pupils of different classes of a given school.

7.4 Linear model

For the linear regression model

$$y_{ijk} = \theta_{ijk} + \varepsilon_{ijk},$$

there are three residuals, as there is variability on three levels. Their variances are denoted by

$$var(\varepsilon_{ijk}) = \sigma^2_\varepsilon,\ var(u_{0jk}) = \sigma^2_{u_0},\ var(v_{00k}) = \sigma^2_{v_{00}}.$$

The total variance between all level-one units now equals $\sigma^2_\varepsilon + \sigma^2_{u_0} + \sigma^2_{v_{00}}$, and the total variance of the level-two units is $\sigma^2_{u_0} + \sigma^2_{v_{00)}}$.

There are several kinds of intraclass correlation coefficient in a three-level model:

Proportion of the total variance from level one:

$$\frac{\sigma^2_\varepsilon}{\sigma^2_\varepsilon + \sigma^2_{u_0} + \sigma^2_{v_{00}}}.$$

Proportion of the total variance from level two:

$$\frac{\sigma^2_{u_0}}{\sigma^2_\varepsilon + \sigma^2_{u_0} + \sigma^2_{v_{00}}}.$$

Proportion of the total variance from level three:

$$\frac{\sigma^2_{v_{00}}}{\sigma^2_\varepsilon + \sigma^2_{u_0} + \sigma^2_{v_{00}}}.$$

Proportion of the total variance from levels one and two:

$$\frac{\sigma^2_\varepsilon + \sigma^2_{u_0}}{\sigma^2_\varepsilon + \sigma^2_{u_0} + \sigma^2_{v_{00}}}.$$

The correlation between different level-1 units (e.g. pupils) of a given level-2 unit (e.g. class) and level-3 unit (e.g. school) is

$$cor\left(y_{ijk}, y_{i'jk} \mid x, z, w\right) = \frac{\sigma^2_{u_0} + \sigma^2_{v_{00}}}{\sigma^2_\varepsilon + \sigma^2_{u_0} + \sigma^2_{v_{00}}},$$

and the correlation between different level-1 units of different level-2 units for a given level-3 unit is

$$cor\left(y_{ijk}, y_{i'j'k} \mid x, z, w\right) = \frac{\sigma^2_{v_{00}}}{\sigma^2_\varepsilon + \sigma^2_{u_0} + \sigma^2_{v_{00}}},$$

so that $cor\left(y_{ijk}, y_{i'jk} \mid x, z, w\right) > cor\left(y_{ijk}, y_{i'j'k} \mid x, z, w\right), i \neq i', j \neq j'$

7.5 Binary Response Model

Discussion of the binary response model focuses on correlations between the different latent responses, e.g. $y^*_{ijk}, y^*_{i'jk}$ and $y^*_{ijk}, y^*_{i'j'k}, i \neq i', j \neq j'$ where

$$y^*_{ijk} = \theta_{ijk} + \varepsilon_{ijk}.$$

For the probit model these correlations are

$$cor\left(y_{ijk}^{*}, y_{i'jk}^{*} \mid x, z, w\right) = \frac{\sigma_{u_0}^2 + \sigma_{v_{00}}^2}{\sigma_{u_0}^2 + \sigma_{v_{00}}^2 + 1},$$

$$cor\left(y_{ijk}^{*}, y_{i'j'k}^{*} \mid x, z, w\right) = \frac{\sigma_{v_{00}}^2}{\sigma_{u_0}^2 + \sigma_{v_{00}}^2 + 1},$$

as $\mathrm{var}(\varepsilon_{ijk}) = 1$.

For the logit model $\mathrm{var}(\varepsilon_{ijk}) = \frac{\pi^2}{3}$ and we replace the 1 in the denominator by $\frac{\pi^2}{3}$.

7.6 Three-Level Generalised Linear Model Likelihood

The 3-level GLM likelihood takes the form

$$\begin{aligned} & L\left(\gamma, \phi, \sigma_{u_0}^2, \sigma_{v_{00}} | \mathbf{y}, \mathbf{x}, \mathbf{z}, \mathbf{w}\right) \\ & = \prod_k \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \prod_j \prod_i g\left(y_{ijk} \mid \theta_{ijk}, \phi\right) f\left(u_{0jk}\right) f\left(v_{00k}\right) du_{0jk} dv_{00k}, \end{aligned}$$

where

$$g\left(y_{ijk} \mid \theta_{ijk}, \phi\right) = \exp\left\{\left[y_{ijk}\theta_{ijk} - b\left(\theta_{ijk}\right)\right]/\phi + c\left(y_{ijk}, \phi\right)\right\},$$

$$\theta_{ijk} = \gamma_{000} + \sum_{p=1}^{P} \gamma_{p00} x_{pijk} + \sum_{q=1}^{Q} \gamma_{0q0} z_{qjk} + \sum_{r=1}^{R} \gamma_{00r} w_{rk} + v_{00k} + u_{0jk},$$

and

$$f\left(u_{0jk}\right) = \frac{1}{\sqrt{2\pi}\sigma_{u_0}} \exp\left(-\frac{u_{0jk}^2}{2\sigma_{u_0}^2}\right),$$

$$f\left(v_{00k}\right) = \frac{1}{\sqrt{2\pi}\sigma_{v_{00}}} \exp\left(-\frac{v_{00k}^2}{2\sigma_{v_{00}}^2}\right).$$

For the linear model we have identity link, Gaussian (normal) error, for the binary model we have one of logit, probit, cloglog link, binomial error, and for the Poisson model we have log link and Poisson error. Sabre evaluates the integral $L\left(\gamma, \phi, \sigma_{u_0}^2, \sigma_{v_{00}} | \mathbf{y}, \mathbf{x}, \mathbf{z}, \mathbf{w}\right)$ for the multilevel GLM model using normal Gaussian quadrature or adaptive Gaussian quadrature (numerical integration).

7.7 Example 3LC2. Binary response model: Guatemalan mothers using prenatal care for their children (1558 mothers in 161 communities)

The data (`guatemala_prenat.dta`) we use in this example are from Rodríguez and Goldman (2001), and are about the use of modern prenatal care. The data set has 2449 observations on children with a binary indicator for whether the mother had prenatal care, there are 25 covariates. The variables include the level-2 mother identifier (mom), the community or cluster (level-3) identifier, a binary indicator of the use of prenatal care for each child and other child-family, and community-level explanatory variables. The explanatory variables are either continuous variables (`pcind81`: proportion indigenous in 1981 and `ssdist`: distance to nearest clinic) or 0-1 dummy variables (all others) representing discrete factors coded using the reference categories. Reference categories are child aged 0-2 years, mother aged <25 years, birth order 1 (eldest child), `ladino` (a Spanish term used to describe various socio-ethnic categories in Central America), mother with no education, husband with no education, husband not working or in unskilled occupation, no modern toilet in household, and no television in the household.

7.7.1 References

G. Rodríguez and N. Goldman (2001) Improved estimation procedures for multilevel models with binary response, Journal of the Royal Statistics Society, Series A, Statistics in Society, Volume 164, Part 2, pages 339-355

7.7.2 Data description for `guatemala_prenat.dta`

Number of observations: 2449
Number of level-2 cases ('mom' = identifier for mothers): 1558
Number of level-3 cases ('cluster' = identifier for communities): 161

7.7.3 Variables

The variables appear in the same order as in Table 3 in G. Rodríguez and N. Goldman (2001) and are:

`kid`: child id (2449 kids)
`mom`: family id (1558 families)
`cluster`: cluster id (161 communities)
`prenat`: 1 if used modern prenatal care, 0 otherwise
`kid3p`: 1 if child aged 3-4 years, 0 otherwise

`mom25p`: 1 if mother aged 25+ years, 0 otherwise
`order23`: 1 if birth order 2-3, 0 otherwise
`order46`: 1 if birth order 4-6, 0 otherwise
`order7p`: 1 if birth order 7+, 0 otherwise
`indnospa`: 1 if indigenous, speaks no Spanish, 0 otherwise
`inspa`: 1 if indgenous, speaks Spanish, 0 otherwise
`momedpri`: 1 if mother's education primary, 0 otherwise
`momedsec`: 1 if mother's education secondary+, 0 otherwise
`husedppri`: 1 if husband's education primary, 0 otherwise
`husedsec`: 1 if husband's education secondary+, 0 otherwise
`huseddk`: 1 if husband's education missing, 0 otherwise
`husprof`: 1 if husband professional, sales, clerical, 0 otherwise
`husagrself`: 1 if husband agricultural self-employed, 0 otherwise
`husagremp`: 1 if husband agricultural employee, 0 otherwise
`husskilled`: 1 if husband skilled service, 0 otherwise
`toilet`: 1 if modern toilet in household, 0 otherwise
`tvnotdaily`: 1 if television not watched daily, 0 otherwise
`tvdaily`: 1 if television watched daily, 0 otherwise
`pcind81`: proportion indigenous in 1981
`ssdist`: distance to nearest clinic

kid	mom	cluster	prenat	kid3p	mom25p	order23	order46	order7p	indnospa	indspa	momedpri	momedsec
2	2	1	1	1	0	0	0	0	0	0	0	1
269	185	36	1	0	0	1	0	0	0	0	1	0
270	186	36	1	0	0	1	0	0	0	0	1	0
271	186	36	1	0	0	1	0	0	0	0	1	0
273	187	36	1	1	0	1	0	0	0	0	1	0
275	188	36	1	1	0	1	0	0	0	0	1	0
276	189	36	1	1	0	1	0	0	0	0	0	1
277	190	36	1	0	1	0	0	1	0	0	1	0
278	190	36	1	1	1	0	1	0	0	0	1	0
279	191	36	1	0	1	0	0	1	0	0	1	0
280	191	36	1	1	1	0	1	0	0	0	1	0
281	192	36	1	0	0	1	0	0	0	0	0	1
299	204	38	1	0	1	0	1	0	0	0	1	0
301	206	38	1	1	1	0	1	0	0	0	1	0
302	207	38	1	0	0	0	0	0	0	0	0	0
358	245	45	1	0	1	1	0	0	0	0	1	0
359	245	45	1	1	0	1	0	0	0	0	1	0
360	246	45	0	0	1	0	0	1	0	0	1	0
361	246	45	1	0	1	0	0	1	0	0	1	0
362	246	45	1	1	1	0	1	0	0	0	1	0
363	247	45	0	0	1	0	0	0	0	0	1	0
364	248	45	1	0	1	0	1	0	0	0	1	0
365	248	45	1	0	1	1	0	0	0	0	1	0
366	249	45	0	0	1	0	1	0	0	0	0	0
367	249	45	0	1	1	0	1	0	0	0	0	0
370	251	45	1	0	1	0	0	1	0	0	0	0
372	252	45	0	1	1	0	0	1	0	0	1	0

First few lines of `guatemala_prenat.dta`

7.7.4 Sabre commands

```
log using guatemala_prenat_s.log, replace
set more off
use guatemala_prenat
#delimit ;
sabre, data kid mom cluster prenat kid3p mom25p order23 order46 order7p
       indnospa indspa momedpri momedsec husedpri husedsec huseddk husprof
       husagrself husagremp husskilled toilet tvnotdaily tvdaily pcind81
       ssdist;
sabre kid mom cluster prenat kid3p mom25p order23 order46 order7p indnospa
      indspa momedpri momedsec husedpri husedsec huseddk husprof husagrself
      husagremp husskilled toilet tvnotdaily tvdaily pcind81 ssdist, read;
#delimit cr
sabre, case first=mom second=cluster
sabre, yvar prenat
sabre, constant cons
#delimit ;
sabre, lfit kid3p mom25p order23 order46 order7p indnospa indspa momedpri
       momedsec husedpri husedsec huseddk husprof husagrself husagremp
       husskilled toilet tvnotdaily tvdaily pcind81 ssdist cons;
#delimit cr
sabre, dis m
sabre, dis e
sabre, mass first=36 second=36
#delimit ;
sabre, fit kid3p mom25p order23 order46 order7p indnospa indspa momedpri
       momedsec husedpri husedsec huseddk husprof husagrself husagremp
       husskilled toilet tvnotdaily tvdaily pcind81 ssdist cons;
#delimit cr
sabre, dis m
sabre, dis e
log close
clear
exit
```

7.7.5 Sabre log file

```
(Standard Homogenous Model)


    Parameter              Estimate         Std. Err.
    ___________________________________________________
    (intercept)             0.71862          0.28529
    kid3p                  -0.20175          0.96881E-01
    mom25p                  0.32066          0.12710
    order23                -0.95341E-01      0.13947
    order46                -0.22862          0.16476
    order7p                -0.18506          0.20115
    indnospa               -0.83905          0.21406
    indspa                 -0.56985          0.16529
    momedpri                0.30643          0.10590
    momedsec                 1.0126          0.28988
    husedpri                0.18462          0.11720
    husedsec                0.67692          0.23795
    huseddk                 0.44553E-02      0.18098
```

```
husprof              -0.32309         0.27313
husagrself           -0.53762         0.23648
husagremp            -0.69955         0.24155
husskilled           -0.36924         0.24374
toilet                0.46521         0.15146
tvnotdaily            0.32393         0.23313
tvdaily               0.46586         0.15248
pcind81              -0.90249         0.20778
ssdist               -0.11460E-01     0.21866E-02


(Random Effects Model)

    Parameter              Estimate         Std. Err.
    ___________________________________________________
    (intercept)             3.5458           1.7266
    kid3p                  -1.0008          0.30401
    mom25p                  1.0253          0.52416
    order23                -0.70703         0.45506
    order46                -0.50441         0.64071
    order7p                -0.97271         0.84425
    indnospa               -5.3249           1.5925
    indspa                 -2.8742           1.0720
    momedpri                1.8261          0.66167
    momedsec                3.9093           1.6070
    husedpri               0.80819          0.68186
    husedsec                3.4292           1.3501
    huseddk                0.57825E-01       1.0320
    husprof                -0.38403          1.5648
    husagrself             -1.7913           1.4116
    husagremp              -2.5822           1.4512
    husskilled             -0.74278          1.4095
    toilet                  1.8823          0.95965
    tvnotdaily              1.4256           1.3987
    tvdaily                 1.4827          0.94111
    pcind81                -4.5496           1.6165
    ssdist                 -0.50489E-01     0.19281E-01
    scale2                  7.0869          0.94235
    scale3                  3.6790          0.61058


    Univariate model
    Standard logit
    Gaussian random effects

    Number of observations            =    2449
    Number of level 2 cases           =    1558
    Number of level 3 cases           =     161

    X-var df           =    22
    Scale df           =     2

    Log likelihood =     -1056.8670     on    2425 residual degrees of freedom
```

7.7.6 Discussion

In this example we use standard Gaussian quadrature with 36 mass points at each level. The results show that there are significant estimated level-2 mother effects (scale2 =7.0869 (s.e. 0.94235)) and level-3 community effects (scale3 =3.6790 (s.e. 0.61058)). The highest correlation in prenatal care is between children of the same mother. Adding the level-2 and level-3 random effects to the linear predictor of a binary response model causes a change in scale of the covariate parameters and a reduction in their significance (relative to the homogeneous model).

For further discussion on 3-level generalised linear models see: Goldstein, (1987), Rabe-Hesketh and Skrondal (2005) and Raudenbush and Bryk (2002)

7.8 Exercises

There are four exercises to accompany this section. Exercise 3LC1 is for a linear model of pupil scores on several questions, where pupils (level 2) are within Schools (level 3). Exercise 3LC2 is for cognative performance on several occasions measured as a binary response of subjects (level 2) within families (level 3). Exercise 3LC3 is for immunization (binary response) of children of mothers (level 2) within comunities (level 3) Exercise 3LC4 is for cancer deaths (Poisson count) of counties within regions (level 2) within nations (level 3).

7.9 References

Goldstein, H., (1987), Multilevel Models in Educational and Social Research, Griffin, London.

Rabe-Hesketh, S., and Skrondal, A., (2005), Multilevel and Longitudinal Modelling using Stata, Stata Press, Stata Corp, College Station, Texas.

Raudenbush, S.W., and Bryk, A.S., (2002), Hierachical Linear Models, Sage, Thousand Oakes, CA.

Chapter 8

Multivariate Two-Level Generalised Linear Mixed Models

8.1 Introduction

We now introduce the superscript r to enable us to distinguish the different models, variates, random effects etc of a multivariate response. There are many examples of this type of data. For instance in a bivariate example the responses could be the wages $\left(y_{ij}^1, r=1\right)$ and trade union membership $\left(y_{ij}^2, r=2\right)$ of an individual j over successive years i. In a different context, Cameron and Trivedi (1988) use various forms of overdispersed Poisson models to study the relationship between type of health insurance and various responses which measure the demand for health care, e.g. number of consultations with a doctor or specialist $\left(y_{ij}^1\right)$ and the number of prescriptions $\left(y_{ij}^2\right)$. An event history example occurs in the modelling of the sequence of months i of job vacancies j, which last until either they are successfully filled $\left(y_{ij}^1\right)$ or withdrawn $\left(y_{ij}^2\right)$ from the market. These data lead to a correlated competing risk model as the firm effects are present in both the filled and lapsed durations, see Andrews et al. at
`http://www.lancs.ac.uk/staff/ecasb/papers/vacdur_economica.pdf`.

A trivariate example is the joint (simultaneous equation) modelling of wages $\left(y_{ij}^1\right)$, training $\left(y_{ij}^2\right)$ and promotion $\left(y_{ij}^3\right)$ of individuals j over time i present in a panel survey such as the British Household Panel Survey (BHPS). Joint modelling of simultaneous responses like allows us to disentangle the direct effects of the different $\left(y_{ij}^r\right)$ on each other from any correlation that occurs in the random effects. Without a multivariate multilevel GLM for complex social process like these we risk inferential errors.

The multivariate GLM is obtained from the univariate GLM (see earlier sections) by specifying the probability of the response $\left(y_{ij}^r\right)$ conditional on the linear predictor and other parameters for each response (r), i.e.

$$g^r\left(y_{ij}^r \mid \theta_{ij}^r, \phi^r\right) = \exp\left\{\left[y_{ij}^r \theta_{ij}^r - b^r\left(\theta_{ij}^r\right)\right]/\phi^r + c^r\left(y_{ij}^r, \phi^r\right)\right\},$$

where ϕ^r is the scale parameter, $b^r\left(\theta_{ij}^r\right)$ is a function that gives the conditional mean $\left(\mu_{ij}^r\right)$ and variance of y_{ij}^r, namely

$$\begin{aligned} E\left[y_{ij}^r \mid \theta_{ij}^r, \phi^r\right] &= \mu_{ij}^r = b^{r'}\left(\theta_{ij}^r\right), \\ Var\left[y_{ij}^r \mid \theta_{ij}^r, \phi^r\right] &= \phi^r b^{r''}\left(\theta_{ij}^r\right), \end{aligned}$$

where the linear predictor $\left(\theta_{ij}^r\right)$ is given by

$$\theta_{ij}^r = \gamma_{00}^r + \sum_{p=1}^{P} \gamma_{p0}^r x_{pij} + \sum_{q=1}^{Q} \gamma_{0q}^r z_{qj} + u_{0j}^r, r = 1, 2, \cdots, R.$$

Both $b^r\left(\theta_{ij}^r\right)$ and $c^r\left(y_{ij}^r, \phi^r\right)$ differ for different members of the exponential family and can be different for different $r, r = 1, 2, \cdots, R$.

8.2 Multivariate 2-Level Generalised Linear Mixed Model Likelihood

We can now write the multivariate 2-level GLM (MGLMM) in general form, i.e.

$$L\left(\gamma, \phi, \Sigma_{u_0} | \mathbf{y}, \mathbf{x}, \mathbf{z}\right) = \prod_j \int_{-\infty}^{\infty} \cdots \int \prod_i \prod_r g^r\left(y_{ij}^r \mid \theta_{ij}^r, \phi^r\right) f\left(\mathbf{u}_{0j}\right) d\mathbf{u}_{0j},$$

where $\gamma = \left[\gamma^1, \gamma^2, ..., \gamma^R\right]$, γ^r has the covariate parameters of the linear predictor θ_{ij}^r, the scale parameters are $\phi = \left[\phi^1, \phi^2, ..., \phi^R\right]$,and $f\left(\mathbf{u}_{0j}\right)$ is a multivariate normal distribution of dimension R with mean zero and variance-covariance structure Σ_{u_0}.

Sabre evaluates the integral $L\left(\gamma, \phi, \Sigma_{u_0} | \mathbf{y}, \mathbf{x}, \mathbf{z}\right)$ in up to 3 dimensions using normal Gaussian quadrature or adaptive Gaussian quadrature (numerical integration).

8.3 Example C6. Bivariate Poisson Model: Number of Visits to the Doctor and Number of Prescriptions

In the 2-level model notation the linear predictor of the bivariate Poisson GLM takes the form

$$\theta_{ij}^r = \gamma_{00}^r + \sum_{p=1}^{P^r} \gamma_{p0}^r x_{pij}^r + \sum_{q=1}^{Q^r} \gamma_{0q}^r z_{qj}^r + u_{0j}^r.$$

The parameters of this model are $\gamma = \left[\gamma^1, \gamma^2\right]$, where γ^r represents the parameters of the linear predictors, plus the two variances $\sigma_{u_0}^1$ and $\sigma_{u_0}^2$ of the random intercepts $\left[u_{0j}^1, u_{0j}^2\right]$ and their correlation is denoted by ρ_{12}.

Cameron and Trivedi (1988) use various forms of overdispersed Poisson models to study the relationship between type of health insurance and various responses which measure the demand for health care, e.g. number of consultations with a doctor or specialist. The data set they use in this analysis is from the Australian Health survey for 1977-1978. In a later work Cameron and Trivedi (1998) estimate a bivariate Poisson model for two of the measures of the demand for health care. We use a version of the Cameron and Trivedi (1988) data set (`visit-prescribe.dta`) for the bivariate model. In this example we only have one pair of responses r(`dvisits`, `prescrib`) for each sampled individual. A copy of the original data set and further details about the variables in `visit-prescribe.dta` can be obtained from
`http://cameron.econ.ucdavis.edu/racd/racddata.html`

The $\sigma_{u_0}^1$, $\sigma_{u_0}^2$ and ρ_{12} can be identified when $i = j = 1$ in bivariate Poisson data. The parameters $\sigma_{u_0}^1$ and $\sigma_{u_0}^2$ are not identifiable when $i = j = 1$ in the binary response-linear model, and to identify these parameters we require $i > 1$.

8.3.1 References

Cameron, A.C., Trivedi, P.K., Milne, F., Piggott, J., (1988) A microeconometric model of the demand for Health Care and Health Insurance in Australia, Review of Economic Studies, 55, 85-106.

Cameron, A.C., Trivedi, P.K (1998), Regression Analysis of Count Data, Econometric Society Monograph No.30, Cambridge University Press.

8.3.2 Data description for `visit-prescribe.dta`

Number of observations (rows): 10380
Number of level-2 cases: 5190

8.3.3 Variables

`sex`: 1 if respondent is female, 0 if male
`age`: respondent's age in years divided by 100
`agesq`: age squared
`income`: respondent's annual income in Australian dollars divided by 1000
`levyplus`: 1 if respondent is covered by private health insurance fund for private patients in public hospital (with doctor of choice), 0 otherwise
`freepoor`: 1 if respondent is covered by government because low income, recent immigrant, unemployed, 0 otherwise
`freerepa`: 1 if respondent is covered free by government because of old-age or disability pension, or because invalid veteran or family of deceased veteran, 0 otherwise
`illness`: number of illnesses in past 2 weeks with 5 or more coded as 5
`actdays`: number of days of reduced activity in past two weeks due to illness or injury
`hscore`: respondent's general health questionnaire score using Goldberg's method, high score indicates poor health
`chcond1`: 1 if respondent has chronic condition(s) but not limited in activity, 0 otherwise
`chcond2`: 1 if respondent has chronic condition(s) and limited in activity, 0 otherwise
`dvisits`: number of consultations with a doctor or specialist in the past 2 weeks
`nondocco`: number of consultations with non-doctor health professionals (chemist, optician, physiotherapist, social worker, district community nurse, chiropodist or chiropractor) in the past 2 weeks
`hospadmi`: number of admissions to a hospital, psychiatric hospital, nursing or convalescent home in the past 12 months (up to 5 or more admissions which is coded as 5)
`hospdays`: number of nights in a hospital, etc. during most recent admission, in past 12 months
`medicine`: total number of prescribed and nonprescribed medications used in past 2 days
`prescrib`: total number of prescribed medications used in past 2 days
`nonpresc`: total number of nonprescribed medications used in past 2 days
`constant`: 1 for all observations
`id`: respondent identifier

ij	r	sex	age	agesq	income	levyplus	freepoor	freerepa	illness	actdays	hscore	chcond1	chcond2	dvisits	nondocco	hospadmi	hospdays	medicine	prescrib	nonpresc	constant	id	y	r1	r2
1	1	1	0.19	0.04	0.55	1	0	0	1	4	1	0	0	1	0	0	0	1	1	0	1	1	1	1	0
1	2	1	0.19	0.04	0.55	1	0	0	1	4	1	0	0	1	0	0	0	1	1	0	1	1	1	0	1
2	1	1	0.19	0.04	0.45	1	0	0	1	2	1	0	0	1	0	0	0	2	1	1	1	2	1	1	0
2	2	1	0.19	0.04	0.45	1	0	0	1	2	1	0	0	1	0	0	0	2	1	1	1	2	1	0	1
3	1	0	0.19	0.04	0.90	0	0	0	3	0	0	0	0	1	0	1	4	2	1	1	1	3	1	1	0
3	2	0	0.19	0.04	0.90	0	0	0	3	0	0	0	0	1	0	1	4	2	1	1	1	3	1	0	1
4	1	0	0.19	0.04	0.15	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	4	1	1	0
4	2	0	0.19	0.04	0.15	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	4	0	0	1
5	1	0	0.19	0.04	0.45	0	0	0	2	5	1	1	0	1	0	0	0	3	1	2	1	5	1	1	0
5	2	0	0.19	0.04	0.45	0	0	0	2	5	1	1	0	1	0	0	0	3	1	2	1	5	1	0	1
6	1	1	0.19	0.04	0.35	0	0	0	5	1	9	1	0	1	0	0	0	1	1	0	1	6	1	1	0
6	2	1	0.19	0.04	0.35	0	0	0	5	1	9	1	0	1	0	0	0	1	1	0	1	6	1	0	1
7	1	1	0.19	0.04	0.55	0	0	0	4	0	2	0	0	1	0	0	0	0	0	0	1	7	1	1	0
7	2	1	0.19	0.04	0.55	0	0	0	4	0	2	0	0	1	0	0	0	0	0	0	1	7	0	0	1
8	1	1	0.19	0.04	0.15	0	0	0	3	0	6	0	0	1	0	0	0	1	1	0	1	8	1	1	0
8	2	1	0.19	0.04	0.15	0	0	0	3	0	6	0	0	1	0	0	0	1	1	0	1	8	1	0	1
9	1	1	0.19	0.04	0.65	1	0	0	2	0	5	0	0	1	0	0	0	1	0	1	1	9	1	1	0
9	2	1	0.19	0.04	0.65	1	0	0	2	0	5	0	0	1	0	0	0	1	0	1	1	9	0	0	1
10	1	0	0.19	0.04	0.15	1	0	0	1	0	0	0	0	1	0	0	0	1	1	0	1	10	1	1	0

First few lines and columns of `racd.dta`

Like Cameron and Trivedi we take `dvisits` and `prescrib` to be the Poisson response variables and model them with a random intercept and a range of explanatory variables. We cross tabulate `dvisits` by `nonprescribe` in the following table.

	nonpresc									
dvisits	0	1	2	3	4	5	6	7	8	total
0	3021	866	180	55	10	4	2	2	1	4141
1	586	145	37	6	4	2	0	2	0	782
2	132	31	9	1	1	0	0	0	0	174
3	23	6	0	0	0	1	0	0	0	30
4	20	2	1	1	0	0	0	0	0	24
5	7	1	1	0	0	0	0	0	0	9
6	10	2	0	0	0	0	0	0	0	12
7	10	1	0	1	0	0	0	0	0	12
8	4	1	0	0	0	0	0	0	0	5
9	1	0	0	0	0	0	0	0	0	1
total	3814	1055	228	64	15	7	2	4	1	5190

Is the assumption of independence between `dvisits` and `nonprescribe` realistic?

Note that in this example $i : 1$ for both responses as we only observe 1 `dvisits` and 1 `prescrib` for each individual.

8.3.4 Sabre commands

```
log using visit-prescribe_s.log, replace
set more off
```

```
use visit-prescribe
#delimit ;
sabre, data ij r sex age agesq income levyplus freepoor freerepa illness
       actdays hscore chcond1 chcond2 dvisits nondocco hospadmi hospdays
       medicine prescrib nonpresc constant id y r1 r2;
sabre ij r sex age agesq income levyplus freepoor freerepa illness actdays
      hscore chcond1 chcond2 dvisits nondocco hospadmi hospdays medicine
      prescrib nonpresc constant id y r1 r2, read;
#delimit cr
sabre, case id
sabre, yvar y
sabre, model b
sabre, rvar r
sabre, family first=p second=p
sabre, constant first=r1 second=r2
sabre, trans r1_sex r1 * sex
sabre, trans r1_age r1 * age
sabre, trans r1_agesq r1 * agesq
sabre, trans r1_income r1 * income
sabre, trans r1_levyplus r1 * levyplus
sabre, trans r1_freepoor r1 * freepoor
sabre, trans r1_freerepa r1 * freerepa
sabre, trans r1_illness r1 * illness
sabre, trans r1_actdays r1 * actdays
sabre, trans r1_hscore r1 * hscore
sabre, trans r1_chcond1 r1 * chcond1
sabre, trans r1_chcond2 r1 * chcond2
sabre, trans r2_sex r2 * sex
sabre, trans r2_age r2 * age
sabre, trans r2_agesq r2 * agesq
sabre, trans r2_income r2 * income
sabre, trans r2_levyplus r2 * levyplus
sabre, trans r2_freepoor r2 * freepoor
sabre, trans r2_freerepa r2 * freerepa
sabre, trans r2_illness r2 * illness
sabre, trans r2_actdays r2 * actdays
sabre, trans r2_hscore r2 * hscore
sabre, trans r2_chcond1 r2 * chcond1
sabre, trans r2_chcond2 r2 * chcond2
sabre, nvar 13
#delimit ;
sabre, lfit r1_sex r1_age r1_agesq r1_income r1_levyplus r1_freepoor
            r1_freerepa r1_illness r1_actdays r1_hscore r1_chcond1 r1_chcond2
            r1
            r2_sex r2_age r2_agesq r2_income r2_levyplus r2_freepoor
            r2_freerepa r2_illness r2_actdays r2_hscore r2_chcond1 r2_chcond2
            r2;
#delimit cr
sabre, dis m
sabre, dis e
sabre, nvar 13
#delimit ;
sabre, fit r1_sex r1_age r1_agesq r1_income r1_levyplus r1_freepoor
           r1_freerepa r1_illness r1_actdays r1_hscore r1_chcond1 r1_chcond2
           r1
           r2_sex r2_age r2_agesq r2_income r2_levyplus r2_freepoor
           r2_freerepa r2_illness r2_actdays r2_hscore r2_chcond1 r2_chcond2
           r2;
#delimit cr
sabre, dis m
sabre, dis e
log close
```

```
clear
exit
```

8.3.5 Sabre log file

```
    Standard Poisson/Poisson

    Number of observations             =   10380

    X-var df           =    26

    Log likelihood =     -8886.3083     on   10354 residual degrees of freedom


    Parameter              Estimate         Std. Err.
    ____________________________________________________
    (intercept).1          -2.2238          0.18982
    sex.1                  0.15688          0.56137E-01
    age.1                   1.0563           1.0008
    agesq.1               -0.84870           1.0778
    income.1              -0.20532          0.88379E-01
    levyplus.1             0.12319          0.71640E-01
    freepoor.1            -0.44006          0.17981
    freerepa.1             0.79798E-01      0.92060E-01
    illness.1              0.18695          0.18281E-01
    actdays.1              0.12685          0.50340E-02
    hscore.1               0.30081E-01      0.10099E-01
    chcond1.1              0.11409          0.66640E-01
    chcond2.1              0.14116          0.83145E-01
    (intercept).2          -2.7412          0.12921
    sex.2                  0.48377          0.36639E-01
    age.2                   2.6497          0.61491
    agesq.2               -0.88778          0.64292
    income.2              -0.44661E-02      0.55766E-01
    levyplus.2             0.28274          0.52278E-01
    freepoor.2            -0.45680E-01      0.12414
    freerepa.2             0.29584          0.59667E-01
    illness.2              0.20112          0.10530E-01
    actdays.2              0.29261E-01      0.36746E-02
    hscore.2               0.20103E-01      0.63664E-02
    chcond1.2              0.77565          0.46130E-01
    chcond2.2               1.0107          0.53895E-01


(Random Effects Model)

 Correlated bivariate model

    Standard Poisson/Poisson
    Gaussian random effects

    Number of observations             =   10380
    Number of cases                    =    5190

    X-var df           =    26
    Scale df           =     3
```

```
Log likelihood =     -8551.2209     on   10351 residual degrees of freedom

Parameter              Estimate        Std. Err.
____________________________________________________
(intercept).1           -2.6694        0.24673
sex.1                    0.27506       0.73571E-01
age.1                   -0.96132        1.3337
agesq.1                  1.4568         1.4522
income.1                -0.11897       0.11257
levyplus.1               0.15202       0.89966E-01
freepoor.1              -0.62151       0.23768
freerepa.1               0.17419       0.12109
illness.1                0.22347       0.25097E-01
actdays.1                0.13872       0.81816E-02
hscore.1                 0.39132E-01   0.14129E-01
chcond1.1                0.15663       0.83179E-01
chcond2.1                0.26404       0.10820
(intercept).2           -2.9069        0.15064
sex.2                    0.57019       0.43558E-01
age.2                    2.0381        0.74431
agesq.2                 -0.19637       0.79300
income.2                 0.32556E-01   0.65766E-01
levyplus.2               0.27330       0.58470E-01
freepoor.2              -0.91061E-01   0.13849
freerepa.2               0.29736       0.69972E-01
illness.2                0.21674       0.13479E-01
actdays.2                0.40222E-01   0.50644E-02
hscore.2                 0.21171E-01   0.81907E-02
chcond1.2                0.77259       0.51285E-01
chcond2.2                1.0204        0.63007E-01
scale1                   0.99674       0.43107E-01
scale2                   0.56067       0.26891E-01
corr                     0.83217       0.52117E-01
```

8.3.6 Discussion

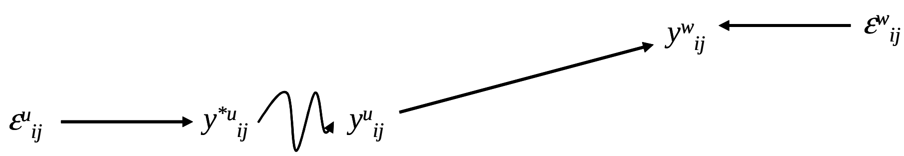
These results show that there is significant overdispersion in both the responses, `dvisits` with `scale1` 0.99674 (s.e. 0.043107) and `prescrib` with `scale2` 0.56067 (s.e. 0.026891), and that these responses are correlated with `corr` 0.83217 (s.e. 0.052117). As expected the standard errors of the estimates of the covariates effects are generally larger in the bivariate GLMM that they are in the homogeneous GLMs.

This shows the different level of overdispersion in the different responses and a large correlation between the random intercepts. If we had not been interested in obtaining the correlation between the responses we could have done a separate analysis of each response and made adjustments to the SEs. This is legitimate here as there are no simultaneous direct effects (e.g. `dvisits` on `precrib`) in this model

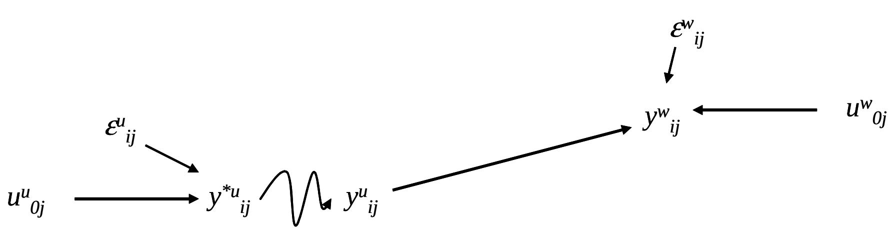
Sabre can model up to 3 different panel responses simultaneously.

8.4 Example L9. Bivariate Linear and Probit Model: Wage and Trade Union Membership

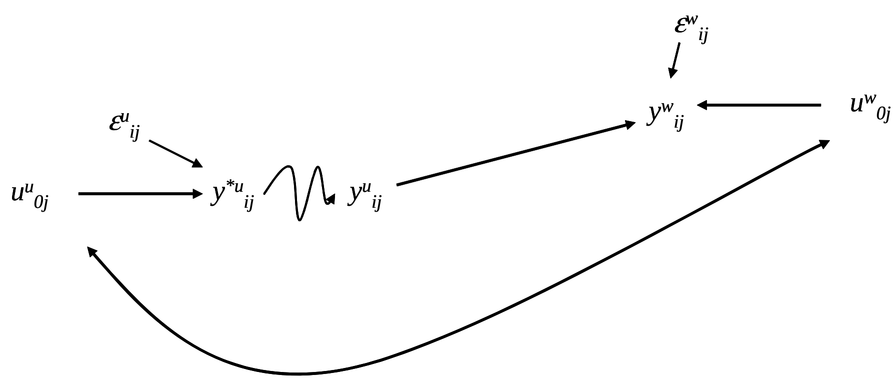
We now illustrate a bivariate multilevel GLM with different link functions. The data we use are versiona (`nls.dta` and `nlswage-union.dta`) of the National Longitudinal Study of Youth (NLSY) data as used in various Stata Manuals (to illustrate the xt commands). The data are for young women who were aged 14-26 in 1968. The women were surveyed each year from 1970 to 1988, except for 1974, 1976, 1979, 1981, 1984 and 1986. We have removed records with missing values on one or more of the response and explanatory variables we want to use in our analysis of the joint determinants of wages and trade union membership. There are 4132 women (idcode) with between 1 and 12 years of observation being in waged employment (i.e. not in full-time education) and earning more than \$1/hour but less than \$700/hour.



The above Figure shows the dependence between trade union membership (y_{ij}^u) and wages (y_{ij}^w). There are no multilevel random effects affecting either wages or trade union membership. The binary response variable trade union membership, $y_{ij}^u = 1, 0$, is based on the latent variable y_{ij}^{*u}. This model can be estimated by any software that estimates basic GLMs.



The above Figure now also shows the dependence between trade union membership and wages. This time there are multilevel random effects affecting both wages and trade union membership. However the multilevel random effects u_{ij}^u and u_{ij}^w are independent, with variances σ_u^2 and σ_w^2 respectively. This model can be estimated by any software that estimates multilevel GLMs by treating the wage and trade union models as independent.



This Figure also shows the dependence between trade union membership and wages, this time there is a correlation ρ_{uw} between the multilevel random effects affecting trade union membership and wages, u_{ij}^{u} and u_{ij}^{w} respectively. This is shown by the curved line linking them together. This model can be estimated by Sabre 5.0 as a bivariate multilevel GLM by allowing for a correlation between the trade union membership wage and wage responses at each wave i of the panel.

How do the results change as the model becomes more comprehensive, especially with regard to the direct effect of trade union membership on wages?

8.4.1 References

Stata Longitudinal/Panel Data, Reference Manual, Release 9, (2005), Stata Press, StataCorp LP, College Station, Texas.

8.4.2 Data description for `nls.dta`

Number of observations: 18995
Number of level-2 cases: 4132

8.4.3 Variables

`ln_wage`: ln(wage/GNP deflator) in a particular year
`black`: 1 if woman is black, 0 otherwise
`msp`: 1 if woman is married and spouse is present, 0 otherwise
`grade`: years of schooling completed (0-18)
`not_smsa`: 1 if woman was living outside a standard metropolitan statistical area (smsa), 0 otherwise
`south`: 1 if woman was living in the South, 0 otherwise
`union`: 1 if woman was a member of a trade union, 0 otherwise

`tenure`: job tenure in years (0-26)
`age`: respondent's age
`age2` : age* age

We will show the differences between a bivariate model and allowing for the correlation between the response sequences. The data displayed below (`nls.dta`), is used for to estimate the separate models for `lnwage` and `union`.

idcode	year	birth_yr	age	race	msp	nev_mar	grade	collgrad	not_smsa	c_city	south	union	ttl_exp	tenure	ln_wage	black	age2	ttl_exp2	tenure2
1	72	51	20	2	1	0	12	0	0	1	0	1	2.26	0.92	1.59	1	400	5.09	0.84
1	77	51	25	2	0	0	12	0	0	1	0	0	3.78	1.50	1.78	1	625	14.26	2.25
1	80	51	28	2	0	0	12	0	0	1	0	1	5.29	1.83	2.55	1	784	28.04	3.36
1	83	51	31	2	0	0	12	0	0	1	0	1	5.29	0.67	2.42	1	961	28.04	0.44
1	85	51	33	2	0	0	12	0	0	1	0	1	7.16	1.92	2.61	1	1089	51.27	3.67
1	87	51	35	2	0	0	12	0	0	0	0	1	8.99	3.92	2.54	1	1225	80.77	15.34
1	88	51	37	2	0	0	12	0	0	0	0	1	10.33	5.33	2.46	1	1369	106.78	28.44
2	71	51	19	2	1	0	12	0	0	1	0	0	0.71	0.25	1.36	1	361	0.51	0.06
2	77	51	25	2	1	0	12	0	0	1	0	1	3.21	2.67	1.73	1	625	10.31	7.11
2	78	51	26	2	1	0	12	0	0	1	0	1	4.21	3.67	1.69	1	676	17.74	13.44
2	80	51	28	2	1	0	12	0	0	1	0	1	6.10	5.58	1.73	1	784	37.16	31.17
2	82	51	30	2	1	0	12	0	0	1	0	1	7.67	7.67	1.81	1	900	58.78	58.78
2	83	51	31	2	1	0	12	0	0	1	0	1	8.58	8.58	1.86	1	961	73.67	73.67
2	85	51	33	2	0	0	12	0	0	1	0	1	10.18	1.83	1.79	1	1089	103.62	3.36
2	87	51	35	2	0	0	12	0	0	1	0	1	12.18	3.75	1.85	1	1225	148.34	14.06
2	88	51	37	2	0	0	12	0	0	1	0	1	13.62	5.25	1.86	1	1369	185.55	27.56
3	71	45	25	2	0	1	12	0	0	1	0	0	3.44	1.42	1.55	1	625	11.85	2.01
3	72	45	26	2	0	1	12	0	0	1	0	0	4.44	2.42	1.61	1	676	19.73	5.84
3	73	45	27	2	0	1	12	0	0	1	0	0	5.38	3.33	1.60	1	729	28.99	11.11
3	77	45	31	2	0	1	12	0	0	1	0	0	6.94	2.42	1.62	1	961	48.20	5.84

First few lines of `nls.dta`

We take `ln_wage` (linear model) and `union` (probit link) to be the response variables and model them with a random intercept and a range of explanatory variables.

Besides allowing for the overdispersion in `ln_wage` and `union`, and correlation between them, the `ln_wage` equation contains `union` as an explanatory variable. We start by estimating separate 2 level models on the sequences of `ln_wage` and `union` from the `nls.dta`, we then use the extended version of the data (`nlswage-union.dta`) needed for the bivariate model. The extended version `nlswage-union.dta` contains the extra variables `r1` and `r2` needed to distinguish the 1st response (`lnwage`) from the second (`union`) in the bivariate model.

8.4.4 Sabre commands

```
log using nlswage-union_s.log, replace
set mem 100m
set more off
use nls
#delimit ;
sabre, data idcode year birth_yr age race msp nev_mar grade collgrad not_smsa
       c_city south union ttl_exp tenure ln_wage black age2 ttl_exp2 tenure2;
sabre idcode year birth_yr age race msp nev_mar grade collgrad not_smsa c_city
      south union ttl_exp tenure ln_wage black age2 ttl_exp2 tenure2, read;
#delimit cr
sabre, case idcode
sabre, yvar ln_wage
sabre, family g
sabre, constant cons
sabre, lfit black msp grade not_smsa south union tenure cons
sabre, dis m
sabre, dis e
sabre, fit black msp grade not_smsa south union tenure cons
sabre, dis m
sabre, dis e
sabre, yvar union
sabre, family b
sabre, link p
sabre, lfit age age2 black msp grade not_smsa south cons
sabre, dis m
sabre, dis e
sabre, fit age age2 black msp grade not_smsa south cons
sabre, dis m
sabre, dis e
clear
use nlswage-union
#delimit ;
sabre, data ij r idcode year birth_yr age race msp nev_mar grade collgrad
       not_smsa c_city south union ttl_exp tenure ln_wage black age2 ttl_exp2
       tenure2 y r1 r2;
sabre ij r idcode year birth_yr age race msp nev_mar grade collgrad not_smsa
      c_city south union ttl_exp tenure ln_wage black age2 ttl_exp2 tenure2 y
      r1 r2, read;
#delimit cr
sabre, case idcode
sabre, yvar y
sabre, model b
sabre, rvar r
sabre, family first=g
sabre, link second=p
sabre, constant first=r1 second=r2
sabre, trans r1_black r1 * black
sabre, trans r1_msp r1 * msp
sabre, trans r1_grade r1 * grade
sabre, trans r1_not_smsa r1 * not_smsa
sabre, trans r1_south r1 * south
sabre, trans r1_union r1 * union
sabre, trans r1_tenure r1 * tenure
sabre, trans r2_age r2 * age
sabre, trans r2_age2 r2 * age2
sabre, trans r2_black r2 * black
sabre, trans r2_msp r2 * msp
sabre, trans r2_grade r2 * grade
sabre, trans r2_not_smsa r2 * not_smsa
```

```
sabre, trans r2_south r2 * south
sabre, nvar 8
#delimit ;
sabre, lfit r1_black r1_msp r1_grade r1_not_smsa r1_south r1_union r1_tenure
            r1
            r2_age r2_age2 r2_black r2_msp r2_grade r2_not_smsa r2_south r2;
#delimit cr
sabre, dis m
sabre, dis e
sabre, nvar 8
#delimit ;
sabre, fit r1_black r1_msp r1_grade r1_not_smsa r1_south r1_union r1_tenure r1
           r2_age r2_age2 r2_black r2_msp r2_grade r2_not_smsa r2_south r2;
#delimit cr
sabre, dis m
sabre, dis e
log close
clear
exit
```

8.4.5 Sabre log file

```
(Standard Homogenous Model)


    Parameter              Estimate         Std. Err.
    ___________________________________________________
    (intercept)             0.82027         0.16614E-01
    black                  -0.10093         0.66150E-02
    msp                     0.50526E-03     0.57363E-02
    grade                   0.69701E-01     0.11861E-02
    not_smsa               -0.18494         0.62495E-02
    south                  -0.80056E-01     0.59837E-02
    tunion                  0.13725         0.66379E-02
    tenure                  0.32222E-01     0.67368E-03
    sigma                   0.37523


(Random Effects Model)


    Parameter              Estimate         Std. Err.
    ___________________________________________________
    (intercept)             0.75217         0.26994E-01
    black                  -0.70564E-01     0.12656E-01
    msp                    -0.12989E-02     0.59885E-02
    grade                   0.72967E-01     0.19959E-02
    not_smsa               -0.14528         0.88414E-02
    south                  -0.73888E-01     0.89322E-02
    tunion                  0.11024         0.65211E-02
    tenure                  0.28481E-01     0.64979E-03
    sigma                   0.26176         0.15024E-02
    scale                   0.27339         0.35702E-02
```

```
    Univariate model
    Standard linear
    Gaussian random effects

    Number of observations             =   18995
    Number of cases                    =    4132

    X-var df           =     8
    Sigma df           =     1
    Scale df           =     1

    Log likelihood =     -4892.5205     on   18985 residual degrees of freedom


(Standard Homogenous Model)


    Parameter              Estimate        Std. Err.
    ___________________________________________________
    (intercept)            -1.3430         0.23760
    age                     0.12788E-01    0.15521E-01
    age2                   -0.10605E-03    0.24659E-03
    black                   0.48206        0.24334E-01
    msp                    -0.20820E-01    0.21552E-01
    grade                   0.31364E-01    0.44733E-02
    not_smsa               -0.75475E-01    0.24045E-01
    south                  -0.49752        0.23085E-01


(Random Effects Model)


    Parameter              Estimate        Std. Err.
    ___________________________________________________
    (intercept)            -2.5916         0.38587
    age                     0.22417E-01    0.23566E-01
    age2                   -0.22314E-03    0.37641E-03
    black                   0.82324        0.68871E-01
    msp                    -0.71011E-01    0.40905E-01
    grade                   0.69085E-01    0.12453E-01
    not_smsa               -0.13402        0.59397E-01
    south                  -0.75488        0.58043E-01
    scale                   1.4571         0.35516E-01


    Univariate model
    Standard probit
    Gaussian random effects

    Number of observations             =   18995
    Number of cases                    =    4132

    X-var df           =     8
    Scale df           =     1

    Log likelihood =     -7647.0998     on   18986 residual degrees of freedom
```

```
(Standard Homogenous Model)

    Parameter              Estimate        Std. Err.
    ____________________________________________________
    (intercept).1          0.82027         0.16614E-01
    black.1               -0.10093         0.66150E-02
    msp.1                  0.50526E-03     0.57363E-02
    grade.1                0.69701E-01     0.11861E-02
    not_smsa.1            -0.18494         0.62495E-02
    south.1               -0.80056E-01     0.59837E-02
    tunion.1               0.13725         0.66379E-02
    tenure.1               0.32222E-01     0.67368E-03
    (intercept).2         -1.3430          0.23760
    black.2                0.48206         0.24334E-01
    msp.2                 -0.20822E-01     0.21552E-01
    grade.2                0.31363E-01     0.44733E-02
    not_smsa.2            -0.75475E-01     0.24045E-01
    south.2               -0.49752         0.23085E-01
    age.2                  0.12788E-01     0.15521E-01
    age2.2                -0.10605E-03     0.24659E-03
    sigma1                 0.37523


(Random Effects Model)

    Parameter              Estimate        Std. Err.
    ____________________________________________________
    (intercept).1          0.75162         0.26753E-01
    black.1               -0.69805E-01     0.12511E-01
    msp.1                 -0.14237E-02     0.59871E-02
    grade.1                0.73275E-01     0.19736E-02
    not_smsa.1            -0.14524         0.88679E-02
    south.1               -0.74533E-01     0.89063E-02
    tunion.1               0.96328E-01     0.70837E-02
    tenure.1               0.28328E-01     0.65261E-03
    (intercept).2         -2.5481          0.38382
    black.2                0.84621         0.69172E-01
    msp.2                 -0.64955E-01     0.41090E-01
    grade.2                0.64562E-01     0.12164E-01
    not_smsa.2            -0.10254         0.58471E-01
    south.2               -0.73260         0.56972E-01
    age.2                  0.20406E-01     0.23558E-01
    age2.2                -0.18467E-03     0.37617E-03
    sigma1                 0.26170         0.15009E-02
    scale1                 0.27466         0.36213E-02
    scale2                  1.4765         0.37284E-01
    corr                   0.11927         0.24144E-01


    Correlated bivariate model

    Standard linear/probit
    Gaussian random effects

    Number of observations           =   37990
    Number of cases                  =    4132

    X-var df          =    16
    Sigma df          =     1
```

```
Scale df          =     3

Log likelihood =     -12529.120     on   37970 residual degrees of freedom
```

8.4.6 Discussion

These last results show the different level of overdispersion in the different responses and a positive correlation between the random intercepts.

The effect of trade union membership in the wage equation changes from 0.11024 (in the model which does allow for the overdispersion of the different responses but not the correlation between them) to 0.09632 which suggests that the effect of the trade union membership on log wages is endogenous.

For further discussion on MGLMMs, see Wooldridge (2002).

8.5 Exercises

There are two MGLMM exercises to accompany this section, namely L9 (bivariate Poisson model) and L10 (joint linear and binary response model).

8.6 References

Wooldridge, J. M. (2002), Econometric Analysis of Cross Section and Panel Data, MIT Press, Cambridge Mass.

Chapter 9

Event History Models

9.1 Introduction

An important type of discrete data occurs with the modelling of the duration to some pre-specified event such as the duration in unemployment from the start of a spell of unemployment until the start of work, the time between shopping trips, or the time to first marriage. This type of discrete data has several important features. For instance, the duration or times to the events of interest are often not observed for all the sampled subjects or individuals. This often happens because the event of interest had not happened by the end of the observation window; when this happens we say that the spell was right censored. This feature is represented in This feature is represented in Figure 9.1.

The Case 4 event has not happened during the period of observation.

The second important feature of social science duration data is that the temporal scale of most social processes is so large (months/years) that it is inappropriate

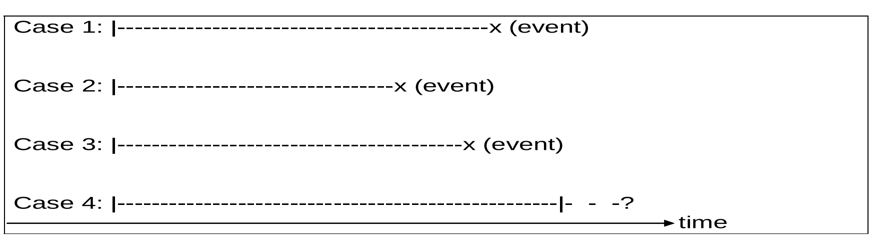


Figure 9.1: Duration Data Schematic

to assume that the explanatory variables remain constant, e.g. in an unemployment spell, the local labour market unemployment rate will vary (at the monthly level) as the local and national economic conditions change. Other explanatory variables like the subject's age change automatically with time.

The third important feature of social science duration data occurs when the observation window cuts into an ongoing spell; this is called left censoring. We will assume throughout that left censoring is non-informative for event history models.

The fourth important feature of duration data is that the spells can be of different types, e.g. the duration of a household in rented accomodation until they move to another rented property could have different characteristics to the duration of a household in rented accommodation until they become owner occupiers. This type of data can be modelled using competing risk models. The theory of competing risks (CR) provides a structure for inference in problems where subjects are exposed to several types of failure. CR models are used in many fields, e.g. in the preparation of life tables for biological populations and in the reliability and safety of engineering systems.

There is a big literature on duration modelling, or what is called survival modelling in medicine. In social science duration data we typically observe a spell over a sequence of intervals, e.g. weeks or months, so we are going to focus on the discrete-time methods. We are not reducing our modelling options by doing this, as durations measured at finer intervals of time such as days, hours, or even seconds can also be written out as a sequence of intervals. We can also group the data by using larger intervals (such as weeks or months) than those at which the durations are measured.

Event history data occur when we observe repeated duration events. If these events are of the same type, e.g. birth intervals, we have a renewal model. When the events can be of different types, e.g. full-time work, part-time work and out of the labour market we have a semi-Markov process. We start by considering a 2-level model for single events (duration model), and then extend this to repeated events of the same kind. We then discuss 3-level models for duration data and end with the multivariate competing risk model.

Various identifiability issues arise in multilevel duration models because of the internal nature of the duration effects on the linear predictor. Identifiability was first discussed for 2-level continuous time models by Elbers and Ridder (1982), and later by Heckman and Singer (1984a, 1984b). These authors show that covariates are needed to identify most 2-level duration models, when the random effect distribution (mixing distribution) has a finite mean (like the Gaussian distribution), the main exception is the Weibull model, which is identified without covariates. These results go through into discrete time models. The identifiability of competing risk models is similar, see Heckman and Honore (1988). Random effect distributions with infinite mean are not covered in this book, for discussion on these see Hougaard (1986a, 1986b).

9.2 Duration Models

Suppose we have a binary indicator y_{ij} for individual j, which takes the value 1 if the spell ends in a particular interval i and 0 otherwise. Then individual j's duration can be viewed as a series of events over consecutive time periods $(i = 1, 2, ..., T_j)$ which can be represented by a binary sequence:

$$\mathbf{y}_j = \left[\mathbf{y}_{1j}, \mathbf{y}_{2j}, ..., \mathbf{y}_{T_j j}\right].$$

If we only observe a single spell for each subject this would be a sequence of 0s, which would end with a 1 if the spell is complete and 0, if it is right censored. We can use the multilevel binary response model notation so that the probability that $y_{ij} = 1$ for individual j at interval i, given that $y_{i'j} = 0, \forall i' < i$ is given by

$$\begin{aligned}\Pr\left(y_{ij} = 1 \mid \theta_{ij}\right) &= 1 - F\left(\theta_{ij}\right) \\ &= \mu_{ij}.\end{aligned}$$

But instead of using the logit or probit link, we use the complementary log log link, which gives

$$\mu_{ij} = 1 - \exp[-\exp\left(\theta_{ij}\right)].$$

This model was derived by Prentice and Gloeckler (1978). The linear predictor takes the form

$$\theta_{ij} = \beta_{0j} + \sum_p \beta_{pj} x_{pij} + k_i,$$

where the k_i are interval-specific constants, the x_{pij} are explanatory variables describing individual and contextual characteristics as before. In survival modelling language the k_i are given by

$$k_i = \log\left\{\Lambda_0\left(t_i\right) - \Lambda_0\left(t_{i-1}\right)\right\},$$

where the $\Lambda_0\left(t_{i-1}\right)$ and $\Lambda_0\left(t_i\right)$ are respectively, the values of the integrated baseline hazard at the start and end of the ith interval.

To help clarify the notation, we give an example of what the data structure would look like for three spells (without covariates). Suppose we had the data:

Subject identifier j	Duration T_j	Censored (1=No, 0=Yes)
1	4	1
2	3	0
3	1	1

Duration data structure

so that e.g. subject 2 has a spell of length 3, which is right censored. Then the data structure we need to model the duration data in discrete time is given in the Table below

Subject Identifier j	Interval i	Response y_{ij}	Interval-specific constants k1	k2	k3	k4
1	1	0	1	0	0	0
1	2	0	0	1	0	0
1	3	0	0	0	1	0
1	4	1	0	0	0	1
2	1	0	1	0	0	0
2	2	0	0	1	0	0
2	3	0	0	0	1	0
3	1	1	1	0	0	0

Duration data structure, modified

To identify the model we need to fix the constant at zero or remove one of the k_i. We often fix the constant at zero.

The likelihood of a subject that is right censored at the end of the T_jth interval is

$$\prod_{i=1}^{T_j} (1 - \mu_{ij}) = \prod_{i=1}^{T_j} \mu_{ij}^{y_{ij}} (1 - \mu_{ij})^{1-y_{ij}} ,$$

where $y_{T_j j} = 0$, while that of a subject whose spell ends without a censoring in the T_jth interval is

$$\mu_{iT_j} \prod_{i=1}^{T_j-1} (1 - \mu_{ij}) = \prod_{i=1}^{T_j} \mu_{ij}^{y_{ij}} (1 - \mu_{ij})^{1-y_{ij}} ,$$

as $y_{T_j j} = 1$.

9.3 Two-level Duration Models

Because the same subject is involved at different intervals we would expect the binary responses y_{ij} and $y_{i'j}$, $i \neq i'$, to be more similar than the responses y_{ij} and $y_{ij'}$, $j \neq j'$. We allow for this similarity with random effects. To allow for the random intercept in the linear predictor

$$\theta_{ij} = \beta_{0j} + \sum_p \beta_{pj} x_{pij} + k_i,$$

we can use multi-level substitutions, with the constraint $\gamma_{00} = 0$, so that

$$\beta_{0j} = \sum_{q=1}^{Q} \gamma_{0q} z_{pj} + u_{0j}, \beta_{pj} = \gamma_{p0},$$

The general model then becomes

$$\theta_{ij} = \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + k_i + u_{0j},$$

and the likelihood becomes

$$L\left(\gamma, k, \phi, \sigma_{u_0}^2 | \mathbf{y}, \mathbf{x}, \mathbf{z}\right) = \prod_j \int_{-\infty}^{+\infty} \prod_i g\left(y_{ij} \mid \theta_{ij}, \phi\right) f\left(u_{0j}\right) du_{0j},$$

with complementary log log link c and binomial error b so that $\phi = 1$, $\mu_{ij} = 1 - \exp\left(-\exp\theta_{ij}\right)$ and

$$g\left(\mathbf{y}_{ij} | \mathbf{x}_{ij}, \mathbf{z}_j, u_{0j}\right) = \mu_{ij}^{y_{ij}} \left(1 - \mu_{ij}\right)^{1-y_{ij}}.$$

Also

$$f\left(u_{0j}\right) = \frac{1}{\sqrt{2\pi}\sigma_{u_0}} \exp\left(-\frac{u_{0j}^2}{2\sigma_{u_0}^2}\right).$$

Sabre evaluates the integral $L\left(\gamma, k, \phi, \sigma_{u_0}^2 | \mathbf{y}, \mathbf{x}, \mathbf{z}\right)$ for this binary response model using numerical quadrature (integration).

9.4 Renewal models

When a subject experiences repeated events of the same type in an observation window we can supply a renewal model. A diagrammatic representation of such data is given by Figure 9.2

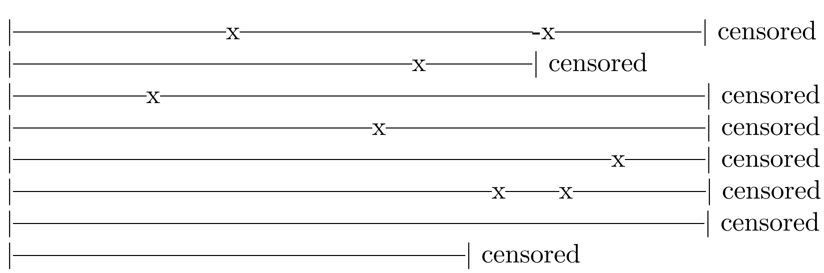


Figure 9.2: Renewal Model Schematic

In the Figure above the subjects that are still present at the end of the observation window have their last event right censored. Two subjects leave the survey before the end of the observation window. Two subjects experience two events each before censoring. Four subjects have one event occurring before they are censored. Two subjects do not experience any events before censoring.

To help clarify the notation, we give an example of what the data structure would look like for 3 subjects observed over 4 intervals (without covariates). Suppose we had

Subject identifier j	Duration T_j	Censored (1=No, 0=Yes)
1	2	1
1	2	0
2	1	1
2	3	0
3	4	0

Renewal data structure

Subject 1 experiences an event after two intervals, followed by two intervals without an event. Subject 2 has an event occurring at the end of interval 1, and is then right censored by the end of interval 4. Subject 3 progresses through all four intervals without experiencing any events.

We now use duration constants (instead of interval constants) to define the duration that occurs in the ith interval. Then the data structure, we need to model the duration data using a binary response GLM , is given by

Subject Identifier j	Interval i	Duration d	Response y_{ij}	Duration-specific constants k1	k2	k3	k4
1	1	1	0	1	0	0	0
1	2	2	1	0	1	0	0
1	3	1	0	1	0	0	0
1	4	2	0	0	0	0	0
2	1	1	1	1	0	0	0
2	2	1	0	1	0	0	0
2	3	2	0	0	0	0	0
2	4	3	0	0	0	1	0
3	1	1	0	1	0	0	0
3	2	2	0	0	1	0	0
3	3	3	0	0	0	1	0
3	4	4	0	0	0	0	1

Renewal data structure, modified

We form the likelihood for the renewal model with the product of $\mu_{ij}^{y_{ij}}\left(1-\mu_{ij}\right)^{1-y_{ij}}$ over the complete sequence. The y_{ij} deal with the occurrence/non-occurrence of the event and the k_d deal with the duration of the spell in the ith interval.

9.5 Example L7. Renewal Model of Residential Mobility

In 1986, the ESRC funded the Social Change and Economic Life Initiative (SCELI). Under this initiative work and life histories were collected for a sample of individuals from 6 different geographical areas in the UK. One of these locations was Rochdale. The data set `roch.dta` contains annual data on male respondents' residential behaviour since entering the labour market. These are residence histories on 348 Rochdale men aged 20 to 60 at the time of the survey. We are going to use these data in the study of the determinants of residential mobility.

9.5.1 Data description for `roch.dta`

Number of observations (rows): 6349
Number of level-2 cases: 348

9.5.2 Variables

`case`: respondent number
`move`: 1 if a residential move occurs during the current year, 0 otherwise
`dur`: number of years since last move
`mbu`: 1 if marriage break-up during the year, 0 otherwise
`fm`: 1 if first marriage during the year, 0 otherwise
`mar`: 1 if married at the beginning of the year, 0 otherwise
`emp`: employment at the beginning of the year (1=self employed; 2=employee; 3=not working)
`age`: (age-30) years
`emp2`: 1 if employment at the beginning of the year is employee; 0 otherwise
`emp3`: 1 if employment at the beginning of the year is not working; 0 otherwise

Note that the variable `dur`, which measures the number of years since the last move is endogenous, i.e. it is internally related to the process of interest.

case	move	dur	mbu	fm	mar	emp	age	emp2	emp3
50004	1	1	0	0	0	2	-13	1	0
50004	0	1	0	0	0	2	-12	1	0
50004	0	2	0	0	0	2	-11	1	0
50004	1	3	0	0	0	2	-10	1	0
50004	0	1	0	0	0	2	-9	1	0
50004	0	2	0	0	0	2	-8	1	0
50004	0	3	0	1	0	3	-7	0	1
50008	0	1	0	0	0	2	-12	1	0
50008	0	2	0	0	0	3	-11	0	1
50008	0	3	0	0	0	3	-10	0	1
50011	0	1	0	0	0	2	-14	1	0
50011	0	2	0	0	0	2	-13	1	0
50011	0	3	0	0	0	2	-12	1	0
50011	0	4	0	0	0	2	-11	1	0
50011	0	5	0	0	0	2	-10	1	0
50011	0	6	0	0	0	2	-9	1	0
50011	0	7	0	0	0	2	-8	1	0
50011	0	8	0	0	0	2	-7	1	0
50011	0	9	0	0	0	2	-6	1	0
50011	0	10	0	0	0	2	-5	1	0
50011	0	11	0	0	0	2	-4	1	0
50011	0	12	0	0	0	2	-3	1	0
50011	0	13	0	0	0	2	-2	1	0
50011	0	14	0	0	0	2	-1	1	0

First few lines of `roch.dta`

We will create quadratic (`age2`) and cubic (`age3`) terms in age to allow more flexibility in modelling this variable (i.e. to allow for a non-linear relationship).

We will then specify the binary response variable (`move`) and fit a cloglog model to the explanatory variables `age dur fm mbu mar emp2 emp3`. Add the `age2` and `age3` effects to this model.

9.5.3 Sabre commands

```
log using roch_s.log, replace
set more off
use roch
sabre, data case move dur mbu fm mar emp age emp2 emp3
sabre case move dur mbu fm mar emp age emp2 emp3, read
sabre, case case
sabre, yvar move
sabre, link c
sabre, constant cons
sabre, trans age2 age * age
sabre, trans age3 age2 * age
sabre, lfit age dur fm mbu mar emp2 emp3 cons
sabre, dis m
sabre, dis e
sabre, fit age dur fm mbu mar emp2 emp3 cons
```

```
sabre, dis m
sabre, dis e
sabre, lfit age dur fm mbu mar emp2 emp3 age2 age3 cons
sabre, dis m
sabre, dis e
sabre, fit age dur fm mbu mar emp2 emp3 age2 age3 cons
sabre, dis m
sabre, dis e
log close
clear
exit
```

9.5.4 Sabre log file

```
(Standard Homogenous Model)

    Parameter              Estimate        Std. Err.
    ___________________________________________________
    (intercept)            -1.4432         0.30719
    age                     0.40490E-01    0.99268E-02
    dur                    -0.19104        0.16430E-01
    fm                      0.66532        0.20423
    mbu                      1.1337        0.60895
    mar                    -0.36649        0.15837
    emp2                   -0.57736E-01    0.28758
    emp3                    0.64292E-01    0.34236


(Random Effects Model)

    Parameter              Estimate        Std. Err.
    ___________________________________________________
    (intercept)            -2.4485         0.38744
    age                     0.20791E-02    0.13319E-01
    dur                    -0.11510        0.20926E-01
    fm                      0.59640        0.21071
    mbu                      1.2865        0.60746
    mar                    -0.52053        0.17935
    emp2                   -0.15696        0.32218
    emp3                   -0.22194E-01    0.37914
    scale                   0.95701        0.12322


    Univariate model
    Standard complementary log-log
    Gaussian random effects

    Number of observations             =    6349
    Number of cases                    =     348

    X-var df           =     8
    Scale df           =     1
```

```
   Log likelihood =     -1092.8370     on    6340 residual degrees of freedom


(Standard Homogenous Model)

   Parameter              Estimate         Std. Err.
   ___________________________________________________
   (intercept)            -1.1106          0.31902
   age                     0.49791E-02     0.18385E-01
   dur                    -0.20439         0.17274E-01
   fm                      0.44789         0.20923
   mbu                     1.0605          0.61012
   mar                    -0.51916         0.15734
   emp2                   -0.41978E-01     0.28697
   emp3                    0.85658E-01     0.34396
   age2                   -0.36339E-02     0.94966E-03
   age3                    0.21321E-03     0.89144E-04


(Random Effects Model)

   Parameter              Estimate         Std. Err.
   ___________________________________________________
   (intercept)            -2.2152          0.40755
   age                    -0.41466E-01     0.20697E-01
   dur                    -0.11896         0.22185E-01
   fm                      0.37503         0.21795
   mbu                     1.2371          0.60712
   mar                    -0.65709         0.18325
   emp2                   -0.17667         0.32416
   emp3                   -0.64809E-01     0.38327
   age2                   -0.27919E-02     0.97393E-03
   age3                    0.25579E-03     0.88150E-04
   scale                   0.95151         0.12350


   Univariate model
   Standard complementary log-log
   Gaussian random effects

   Number of observations             =    6349
   Number of cases                    =     348

   X-var df           =    10
   Scale df           =     1

   Log likelihood =     -1085.6462     on    6338 residual degrees of freedom
```

9.5.5 Discussion

The addition of variables `age2` {coefficient -0.0027919 (s.e. 0.00097393)} and `age3` {coefficient 0.00025579 (s.e. 0.000088150)} to the model has significantly reduced the log likelihood. Age clearly has a complicated realtionship with the probability of moving. The duration effect `dur` has coeffcient -0.11896 (s.e. 0.022185), which suggests that the respondent is less likely to move the longer they stay in their current home. The level-2 random effect is very significant, it has the parameter `scale` and takes the value 0.95151 (s.e. 0.12350).

9.5.6 Exercise

Exercise L11 is a renewal model exercise on repeated times to angina pectoris.

9.6 Three-level Duration Models

We can also apply 3-level event history models to duration data. The binary response variable, which now needs to acknowledge the extra level, is denoted by y_{ijk}, e.g. referring to the modelling of firm vacancies, where $y_{ijk} = 1$ if the vacancy is filled in interval i of vacancy j of firm k and $y_{ijk} = 0$ otherwise. We would expect that the duration of vacancies of a particular firm to be more similar than the duration of vacancies of different firms. We would also expect that the binary responses y_{ijk} and $y_{i'jk}$ to be more similar than those of different j.

9.6.1 Exercises

The Exercise 3LC5 is a 3 level vacancy duration model with 1736 vacancies in 515 firms.

9.7 Competing Risk Models

The theory of competing risks (CR) provides a structure for inference in problems where subjects are exposed to several types of event. We earlier gave the example of a household in rented accommodation, moving to different rented accommodation or becoming an owner occupier (2 possible types of ending). An example in the labour market context is given by a spell of unemployment ending in employment in a skilled, semi-skilled or unskilled occupation (3 possible types of ending). Because the same subjects are exposed to the possibility of different types of events occurring, we would expect that in addition to the probability of a particular event occurring at a given interval being correlated with the probability of that event occurring at another interval, the probability of the different events occurring are also correlated.

The Figure 9.3 shows failure/death due to two failure mechanisms A and B. Three observations are terminated by events of type A. Events of type B occur for three further subjects. Two observations are censored.

```
|--------------------------------------------------x A
|---------------------------------------x B
|-----------------------------------------------------------------| censored
|-----------------------------------x A
|-----------------------------------------------------------x B
|-----------------------------------------------xB
|------------------------------------------------------------xA
|---------------------------------------------| censored
```

Figure 9.3: Failure/Death Due To Two Failure Mechanisms Schematic

```
|--------------------------------------------------x A
|----------------------------------------|censored
|-----------------------------------------------------------------| censored
|-----------------------------------x A
|------------------------------------------------------------| censored
|------------------------------------------------|censored
|------------------------------------------------------------xA
|---------------------------------------------| censored
```

Figure 9.4: Data for the model for failure due to mechanism A

To model failure type A.Define an event as a time when a failure of type A occurs, and treat all other observations as censored, ie. if a failure of type B occurs at time $t1$, this is regarded as a censoring at time $t1$ as far as process A is concerned, as a failure of type A has not yet occurred by time $t1$.

Analyse replications of the data for each failure type.

In Table 9.1 we present some sample competing risk data of the times to two events (A & B) for 3 subjects. Subject 1 has an event of type A occurring by the end of interval 2. Subject 2 is censored at the end of interval 2 without an event occurring. Subject 3 experiences an event of type B by the end of interval 4.

```
|--------------------------------------------------|censored
|---------------------------------------x B
|-----------------------------------------------------------------| censored
|-----------------------------------|censored
|-----------------------------------------------------------x B
|-----------------------------------------------xB
|------------------------------------------------------------|censored
|---------------------------------------------| censored
```

Figure 9.5: Data for the model for failure due to mechanism B

Subject identifier j	Duration T_j	Event (1=A,2=B)	Censored (1=No, 0=Yes)
1	2	1	1
1	2	2	0
2	1	1	0
2	1	2	0
3	4	1	0
3	4	2	1

Table 9.1: Competing risk data structure

Subject Identifier j	Interval i	Duration d	Response y_{ij}	Event 1=A,2=B	Duration-specific constants k1	k2	k3	k4
1	1	1	0	1	1	0	0	0
1	2	2	1	1	0	1	0	0
1	1	1	0	2	1	0	0	0
1	2	2	0	2	0	1	0	0
2	1	1	0	1	1	0	0	0
2	1	1	0	2	1	0	0	0
3	1	1	0	1	1	0	0	0
3	2	2	0	1	0	1	0	0
3	3	3	0	1	0	0	1	0
3	4	4	0	1	0	0	0	1
3	1	1	0	2	1	0	0	0
3	2	2	0	2	0	1	0	0
3	3	3	0	2	0	0	1	0
3	4	4	1	2	0	0	0	1

Table 9.2: Competing risk data structure, modified

9.8 Likelihood

$$L\left(\gamma, \mathbf{k}, \phi, \Sigma_{u_0} | \mathbf{y}, \mathbf{x}, \mathbf{z}\right) = \prod_j \int_{-\infty}^{\infty} \cdots \int \prod_i \prod_r g^r\left(y_{ij}^r \mid \theta_{ij}^r, \phi^r\right) f\left(\mathbf{u}_{0j}\right) d\mathbf{u}_{0j},$$

with cloglog link c and binomial error b so that $\phi^r = 1,\ \mu_{ij}^r = 1 - \exp\left(-\exp\theta_{ij}^r\right)$,

$$g^r\left(y_{ij}^r | \theta_{ij}^r, \phi^r\right) = \left(\mu_{ij}^r\right)^{y_{ij}^r}\left(1 - \mu_{ij}^r\right)^{1-y_{ij}^r},$$

$$\theta_{ij}^r = \sum_{p=1}^{P} \gamma_{p0}^r x_{pij} + \sum_{q=1}^{Q} \gamma_{0q}^r z_{qj} + k_i^r + u_{0j}^r,$$

and where $\gamma = \left[\gamma^1, \gamma^2, ..., \gamma^R\right]$, γ^r has the covariate parameters of the linear predictor θ_{ij}^r, $\mathbf{k} = \left[\mathbf{k}^1, \mathbf{k}^2, ..., \mathbf{k}^R\right]$, and $f\left(\mathbf{u}_{0j}\right)$ is a multivariate normal distribution of dimension R with mean zero and variance-covariance structure Σ_{u_0}.Sabre evaluates the integral $L\left(\gamma, \mathbf{k}, \phi, \Sigma_{u_0} | \mathbf{y}, \mathbf{x}, \mathbf{z}\right)$ using normal Gaussian quadrature or adaptive Gaussian quadrature (numerical integration).

9.9 Example L8. Correlated Competing Risk Model of Filled and Lapsed Vacancies

This example is from a study of the determinants of employer search in the UK using duration modelling techniques. It involves modelling a job vacancy duration until either it is successfully filled or withdrawn from the market. For further details, see Andrews et al (2005). The model has a filled random effect for the filled sequence and a lapsed random effect for the lapsed sequence. Rather than treat the 'filled' and 'lapsed' response sequences as if they were independent from each other, we allow for a correlation between the random effects. There are 7,234 filled vacancies and 5,606 lapsed vacancies.

For each type of risk we used a Weibull baseline hazard, i.e. with `log t` in the linear predictor of the complementary log log links and for simplicity the same 6 covariates. The combined dataset (`fillap-c.dta`), has 22,682 observations, each of the 2,374 vacancies being represented twice, with each sequence of vacancy responses ending in a 1 at the point where the vacancy is filled for a 'filled' risk, the 'lapsed' risk is right censored at this point and vice versa for a 'lapsed' risk.

There are a range of questions that the substantive researcher might be interested in. These include: what is the significance and magnitude of the the random effects of each risk (if any) and what is the sign and magnitude of the correlation between the risks? Would you expect this correlation to be negative or positive? We may also be interested in comparing the results of the bivariate model with those of the uncorrelated model, as the results may change as the model becomes more comprehensive, especially with regard to the inference on the covariates.

9.9.1 References

Andrews, M.J., Bradley, S., Stott, D., Upward, R., (2005), Successful employer search? An empirical analysis of vacancy duration using micro data, see http://www.lancs.ac.uk/staff/ecasb/papers/vacdur_economica.pdf.

9.9.2 Data description for `fillap-c.dta`

Number of observations: 22682
Number of level-2 cases: 2374

9.9.3 Variables

`r`: response index, 1 for filled, 2 for lapsed
`vacnum`: vacancy reference number

`hwage`: hourly wage
`noemps1`: 1 if <= 10 employees, 0 otherwise
`noemps2`: 1 if 11-30 employees, 0 otherwise
`noemps3`: 1 if 31-100 employees, 0 otherwise
`noemps4`: 1 if > 100 employees, 0 otherwise
`y`: response variable, 1 if vacancy filled, 0 otherwise in a particular week
`nonman`: 1 if a non-manual vacancy, 0 otherwise
`skilled`: 1 if a skilled occupation, 0 otherwise
`logt`: log vacancy duration in weeks
`r1`: 1 if first response filled, 0 otherwise
`r2`: 1 if second response lapsed, 0 otherwsie

r	vacnum	hwage	noemps2	noemps3	noemps4	y	nonman	skilled	logt	r1	r2
1	2838	1.05	0	0	0	0	1	1	0.0000	1	0
1	2838	1.05	0	0	0	0	1	1	0.6931	1	0
1	2838	1.05	0	0	0	0	1	1	1.0986	1	0
2	2838	1.05	0	0	0	0	1	1	0.0000	0	1
2	2838	1.05	0	0	0	0	1	1	0.6931	0	1
2	2838	1.05	0	0	0	1	1	1	1.0986	0	1
1	2843	1.23	0	0	0	1	1	1	0.0000	1	0
2	2843	1.23	0	0	0	0	1	1	0.0000	0	1
1	2846	1.25	0	0	0	0	0	0	0.0000	1	0
1	2846	1.25	0	0	0	0	0	0	0.6931	1	0
1	2846	1.25	0	0	0	1	0	0	1.0986	1	0
2	2846	1.25	0	0	0	0	0	0	0.0000	0	1
2	2846	1.25	0	0	0	0	0	0	0.6931	0	1
2	2846	1.25	0	0	0	0	0	0	1.0986	0	1
1	2847	2.52	0	0	1	0	1	1	0.0000	1	0
1	2847	2.52	0	0	1	0	1	1	0.6931	1	0
1	2847	2.52	0	0	1	0	1	1	1.0986	1	0
1	2847	2.52	0	0	1	0	1	1	1.3863	1	0
1	2847	2.52	0	0	1	0	1	1	1.6094	1	0
1	2847	2.52	0	0	1	1	1	1	1.7918	1	0
2	2847	2.52	0	0	1	0	1	1	0.0000	0	1
2	2847	2.52	0	0	1	0	1	1	0.6931	0	1
2	2847	2.52	0	0	1	0	1	1	1.0986	0	1

First few lines of `filap-c.dta`

9.9.4 Sabre commands

```
log using filled-lapsed_s.log, replace
set mem 200m
set more off
use fillap-c
sabre, data r vacnum hwage noemps2 noemps3 noemps4 y nonman skilled logt r1 r2
sabre r vacnum hwage noemps2 noemps3 noemps4 y nonman skilled logt r1 r2, read
sabre, case vacnum
sabre, yvar y
sabre, model b
sabre, rvar r
sabre, link first=c second=c
sabre, constant first=r1 second=r2
```

```
sabre, trans r1_logt r1 * logt
sabre, trans r1_noemps2 r1 * noemps2
sabre, trans r1_noemps3 r1 * noemps3
sabre, trans r1_noemps4 r1 * noemps4
sabre, trans r1_hwage r1 * hwage
sabre, trans r1_nonman r1 * nonman
sabre, trans r1_skilled r1 * skilled
sabre, trans r2_logt r2 * logt
sabre, trans r2_noemps2 r2 * noemps2
sabre, trans r2_noemps3 r2 * noemps3
sabre, trans r2_noemps4 r2 * noemps4
sabre, trans r2_hwage r2 * hwage
sabre, trans r2_nonman r2 * nonman
sabre, trans r2_skilled r2 * skilled
sabre, nvar 8
#delimit ;
sabre, lfit r1 r1_logt r1_noemps2 r1_noemps3 r1_noemps4 r1_hwage r1_nonman
            r1_skilled
            r2 r2_logt r2_noemps2 r2_noemps3 r2_noemps4 r2_hwage r2_nonman
            r2_skilled;
#delimit cr
sabre, dis m
sabre, dis e
sabre, corr n
sabre, mass first=32 second=32
sabre, nvar 8
#delimit ;
sabre, fit r1 r1_logt r1_noemps2 r1_noemps3 r1_noemps4 r1_hwage r1_nonman
           r1_skilled
           r2 r2_logt r2_noemps2 r2_noemps3 r2_noemps4 r2_hwage r2_nonman
           r2_skilled;
#delimit cr
sabre, dis m
sabre, dis e
sabre, corr y
sabre, mass first=32 second=32
sabre, nvar 8
#delimit ;
sabre, fit r1 r1_logt r1_noemps2 r1_noemps3 r1_noemps4 r1_hwage r1_nonman
           r1_skilled
           r2 r2_logt r2_noemps2 r2_noemps3 r2_noemps4 r2_hwage r2_nonman
           r2_skilled;
#delimit cr
sabre, dis m
sabre, dis e
log close
clear
exit
```

9.9.5 Sabre log file

```
    Standard complementary log-log/complementary log-log

    Number of observations             =   22682

    X-var df          =    16
```

```
Log likelihood =     -7294.3634     on   22666 residual degrees of freedom



Parameter              Estimate        Std. Err.
___________________________________________________
r1                    -0.71874         0.10059
r1_logt               -0.54650         0.31561E-01
r1_noemps2             0.10279E-01     0.70765E-01
r1_noemps3             0.10332         0.77017E-01
r1_noemps4            -0.24391         0.86725E-01
r1_hwage              -0.41088         0.77695E-01
r1_nonman             -0.38542E-01     0.62045E-01
r1_skilled            -0.19629         0.61639E-01
r2                     -2.1043         0.11009
r2_logt                0.10615         0.32886E-01
r2_noemps2            -0.17174         0.84481E-01
r2_noemps3            -0.39823         0.10040
r2_noemps4            -0.30067         0.91783E-01
r2_hwage              -0.16668         0.77912E-01
r2_nonman              0.10625         0.72740E-01
r2_skilled            -0.13553         0.72238E-01


Independent bivariate model

Standard complementary log-log/complementary log-log
Gaussian random effects

Number of observations            =   22682
Number of cases                   =    2374

X-var df          =    16
Scale df          =     2

Log likelihood =     -7287.7119     on   22664 residual degrees of freedom



Parameter              Estimate        Std. Err.
___________________________________________________
r1                    -0.73882         0.12176
r1_logt               -0.33126         0.11457
r1_noemps2            -0.62864E-02     0.84908E-01
r1_noemps3             0.95274E-01     0.92602E-01
r1_noemps4            -0.24901         0.10348
r1_hwage              -0.50311         0.10228
r1_nonman             -0.91256E-01     0.78260E-01
r1_skilled            -0.24494         0.77469E-01
r2                     -2.3474         0.19325
r2_logt                0.39002         0.14721
r2_noemps2            -0.21549         0.10585
r2_noemps3            -0.49738         0.13083
r2_noemps4            -0.33570         0.11693
r2_hwage              -0.21624         0.10120
r2_nonman              0.88611E-01     0.89750E-01
r2_skilled            -0.18809         0.91930E-01
scale1                 0.71227         0.20805
scale2                 0.76498         0.23191


Correlated bivariate model
```

```
Standard complementary log-log/complementary log-log
Gaussian random effects

Number of observations            =   22682
Number of cases                   =    2374

X-var df          =    16
Scale df          =     3

Log likelihood =     -7217.7072     on   22663 residual degrees of freedom


Parameter             Estimate        Std. Err.
____________________________________________________
r1                   -0.96329         0.15666
r1_logt              -0.35523         0.87898E-01
r1_noemps2            0.37481E-01     0.10638
r1_noemps3            0.18021         0.11994
r1_noemps4           -0.23653         0.13044
r1_hwage             -0.53793         0.12288
r1_nonman            -0.10936         0.95910E-01
r1_skilled           -0.26235         0.96725E-01
r2                    -7.6478          1.5206
r2_logt                2.7385         0.72471
r2_noemps2           -0.75970         0.38966
r2_noemps3            -1.6889         0.51056
r2_noemps4            -1.0762         0.44983
r2_hwage             -0.26480         0.39838
r2_nonman             0.47016         0.32326
r2_skilled           -0.36773         0.33192
scale1                 1.2887         0.16222
scale2                 5.2516          1.1412
corr                 -0.89264         0.35399E-01
```

9.9.6 Discussion

These results show what happens when we first add vacancy specific random effects, there is a change in likelihood of

$$-2(-7294.3634 - (-7287.7119)) = 13.303,$$

over the homogeneous model. When we allow for a correlation between the random effects of the filled and lapsed durations there is a change in likelhood of

$$-2(-7287.7119 - (-7217.7072)) = 140.01,$$

over the model that assumes indepdence between the filled and lapsed exits from a vacancy. These last results show the different level of overdispersion in the different responses and a negative correlation between the random effects of the two risks. This may be expected, as a filled vacancy can not lapse and vice versa. The random effect of the filled vacancies has a standard deviation of 1.2887 (s.e. 0.16222),and that of the lapsed vacancies is a lot larger at 5.2516 (s.e. 1.1412), their correlation is -0.89264 (s.e. 0.035399).

It should be noticed that the inference on duration effects from the model that assumes independence between the random effcts of the filled and lapsed durations are quite different to those that allow for a correlation. For instance `r2_logt` 39002 (s.e. 0.14721),becomes 2.7385 (s.e. 0.72471).in the correlated model. The value of the coefficient on `r2_logt` suggests that the longer a vacancy goes unfilled the longer it is likley to be unfilled. Differences also occur for the firm size effects (r2_`noemps)` which are over 2 times bigger in the correlated model.

9.9.7 Exercises

Exercises L12 is for a multivariate competing risk model.

9.10 References

Elbers, C., and Ridder, G., (1982), True and Spurious Duration Dependence: The Identifiability of the Proportional Hazards Model, Review of Economics Studies, 49, 402-410.

Heckman, J.J., and Singer, B., (1984a), Econometric Duration Analysis, Journal of Econometrics, 24, 63-132.

Heckman, J.J., and Singer, B., (1984b), The Identifiability of the Proportional Hazards Model, Review of Economics Studies, 51, 231-241.

Heckman, J.J., and Honore, B.E., (1988), The Identifiability of the Competing Risks Model, Biometrika, 76, 325-330.

Hougaard, P., (1986a), Survival Models for Heterogenous Populations Derived from Stable Distributions, Biometrika, 73, 387-396.

Hougaard, P., (1986b), A Class of Multivariate Failure Time Distributions, Biometrika, 73, 671-678.

Chapter 10

Stayers, Non-susceptibles and Endpoints

10.1 Introduction

There are several empirical contexts in which a subset of the population might behave differently to those that follow the proposed GLMM. For instance, in a migration study, we could observe a large group of respondents who do not move outside the study region over the study period. These observed non-migrators could be made up of two distinct groups: those that consider migrating, but are not observed to do so; and those that would never ever consider migrating (the stayers). This phenomenon can occur in various contexts, e.g. zero-inflated Poisson Model (Green,1994 and Lambert, 1992); the mover-stayer model (Goodman, 1961) and in the competing risk context, where some individuals are not vulnerable to an exit condition, e.g. few unemployed males will seek part-time work. In biometric research, these 'stayers' are often referred to as non-susceptibles.

It has often been noted that the goodness-of-fit of mixture models like GLMMs can be improved by adding a spike to the parametric distribution for the random effects to represent stayers, explicitly resulting in a 'spiked distribution' (Singer and Spillerman, 1976). Non-parametric representations of the random effects distribution, e.g. Heckman and Singer(1984), Davies and Crouchley (1986) can have the flexibility to accommodate stayers. However, non parametric random effects distributions can require a lot of parameters (mass point locations and probabilities), while spiked distributions are generally more parsimonious.

Sabre assumes a Gaussian or normal probability distribution for the random effects with mean zero and standard deviation to be estimated from the data, see Figure 10.1.

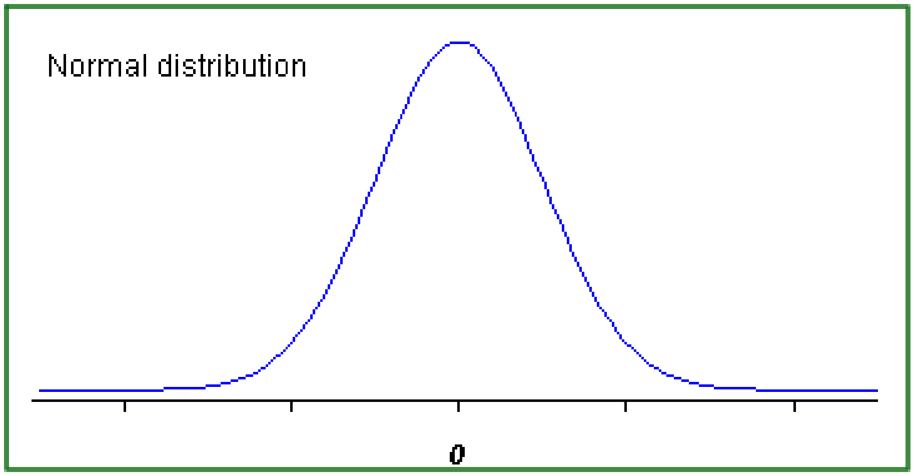


Figure 10.1. The Normal Distribution

This distribution is approximated by a number of mass (or quadrature) points with specified probabilities at given locations. This is illustrated by the solid vertical lines in Figure 10.2. Increasing the number of quadrature points, increases the accuracy of the computation at the expense of computer time.

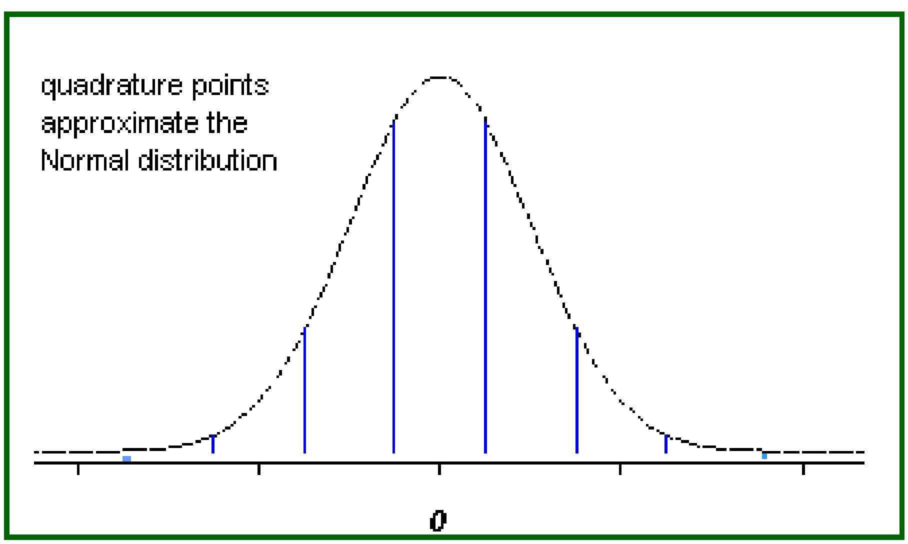


Figure 10.2. Quadrature Points Approximate the Normal Distribution

To compensate for the limitations of the Gaussian distribution for the random effects (i.e. tending to zero too quickly at the extremes), Sabre has the flexibility to supplement the quadrature points with endpoints (i.e. delta functions at plus and/or minus infinity) whose probabilities can be estimated from the data, see Figure 10.3. This flexibility may be needed when modelling binary data.

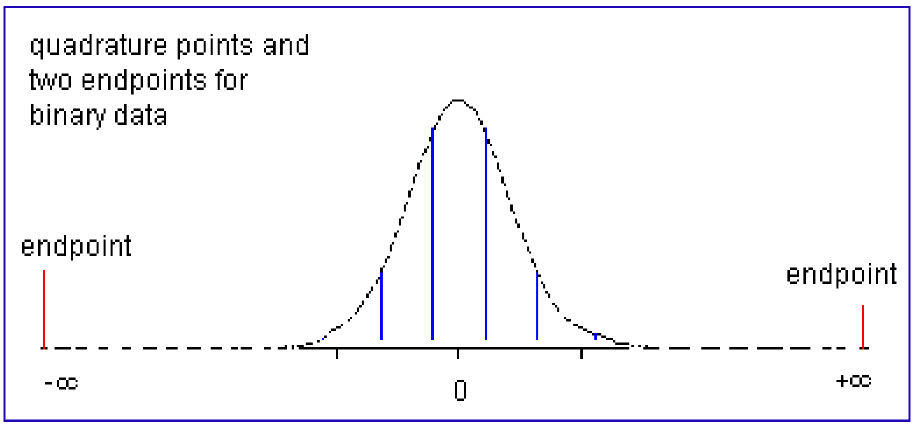


Figure 10.3. Quadrature with Left and Right Endpoints

With the Poisson model, a single left endpoint at minus infinity, see Figure 10.4, allows for extra zeros.

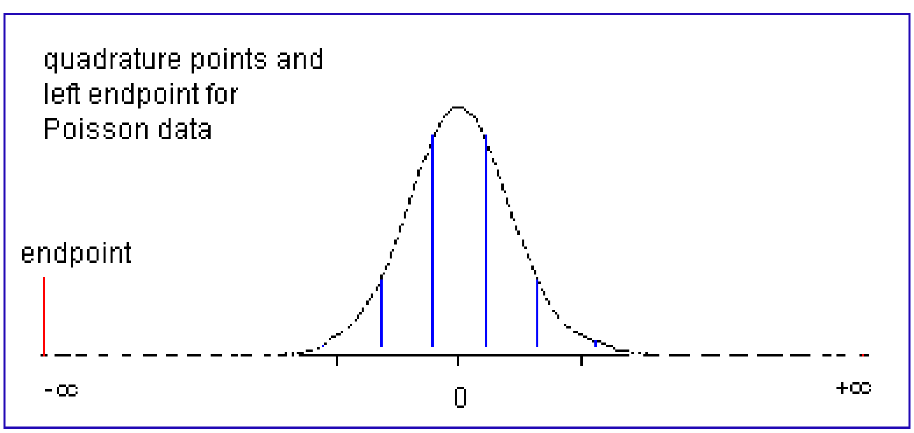


Figure 10.4. Quadrature Points and Left Endpoint

10.2 Likelihood with Endpoints

To allow for stayers in GLMMs, we need to extend our notation. Let the two types of 'stayer' be denoted by S_r and S_l for the right (plus infinity) and left (minus infinity) spikes and let the probability of these events be $\Pr[S_r]$ and $\Pr[S_l]$.

In a binary response 2-level GLMM, let T_j be the length of the observed sequence, and $\Sigma_j = \sum_i y_{ij}$, where y_{ij} is the binary response of individual j at occasion i. Let $S_l = [0, 0, ..., 0]$ represent a sequence without any moves, and let $S_r = [1, 1, ..., 1]$ represent a sequence with moves at every point. The likelihood of the binary response GLMM with endpoints takes the form

$$L\left(\gamma, \phi, \sigma_{u_0}^2 | \mathbf{y}, \mathbf{x}, \mathbf{z}\right) = \prod_j \left\{ \begin{array}{c} \Pr[S_l] .0^{\Sigma_j} + \Pr[S_r] .0^{T_j - \Sigma_j} + \\ (1 - \Pr[S_l] - \Pr[S_r]) \int \prod_i g\left(y_{ij} \mid \theta_{ij}, \phi\right) f\left(u_{0j}\right) du_{0j} \end{array} \right\},$$

where

$$g\left(y_{ij} \mid \theta_{ij}, \phi\right) = \exp\left\{\left[y_{ij}\theta_{ij} - b\left(\theta_{ij}\right)\right]/\phi + c\left(y_{ij}, \phi\right)\right\},$$

$$\theta_{ij} = \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + u_{0j},$$

and

$$f\left(u_{0j}\right) = \frac{1}{\sqrt{2\pi}\sigma_{u_0}} \exp\left(-\frac{u_{0j}^2}{2\sigma_{u_0}^2}\right),$$

as before. We parameterise $\Pr[S_l]$ and $\Pr[S_r]$as

$$\Pr[S_l] = \frac{l}{1+l},$$
$$\Pr[S_r] = \frac{r}{1+r},$$

where $l, r > 0$.

In a zero-inflated Poisson GLMM, $S_l = [0, 0, ..., 0]$ represents a sequence with zero counts at every point. There is no S_r, so that $\Pr[S_r] = 0$. Then the above likelihood simplifies to:

$$L\left(\gamma, \phi, \sigma_{u_0}^2 | \mathbf{y}, \mathbf{x}, \mathbf{z}\right) = \prod_j \left\{ \begin{array}{c} \Pr[S_l] .0^{\Sigma_j} + \\ (1 - \Pr[S_l]) \int \prod_i g\left(y_{ij} \mid \theta_{ij}, \phi\right) f\left(u_{0j}\right) du_{0j} \end{array} \right\}.$$

The binary and Poisson models can be extended in two ways: (1) to allow for between individual (j) variation in the probability of being a stayer, we can make $\Pr[S_l]$ (and $\Pr[S_r]$) a function of time-constant covariates and write $\Pr[S_{lj}]$ (and $\Pr[S_{rj}]$); (2) to allow $\Pr[S_l]$ to vary over response occasions (i) as well as between individuals (j) we can write $\Pr[S_{lij}]$. However, these extensions have not yet been implemented in Sabre.

10.3 End-points: Poisson and Binary Response Examples

Both of these examples are concerned with individuals' migration histories within Great Britain, where migration is a residential move between two counties.

The data we use are derived from a large retrospective survey of life and work histories carried out in 1986 under the Social Change and Economic Life Initiative (SCELI), funded by the ESRC. The data were therefore not specifically collected for the study of migration, but were drawn from an existing data set which includes information on where individuals had lived all their working lives. Temporary moves of a few months duration do not imply commitment to a new area and are not regarded as migration. Migration data are therefore recorded on an annual basis.

The respondents were aged 20 to 60 and lived in the travel-to-work area of Rochdale, just to the north of Manchester, UK. (Rochdale was one of six localities chosen for their contrasting experience of recent economic change.) As the analysis is concerned with internal migration within Great Britain, individuals who had lived abroad during their working lives are excluded from the data set. For simplicity, we ignore the complications due to differential pushes and pulls of different regions in the following Poisson and binary response models of migration behaviour.

10.3.1 Poisson Model

For each individual, we have summed the number of annual migrations recorded in the survey, to produce one line of information. This information is contained in `rochmigx.dta`. Table 10.1 summarizes the observed migration frequencies for the 348 respondents in the sample. As the individuals ranged in age from 20 to 60, they have varying lengths of migration history.

Number of moves	0	1	2	3	4	5	>=6
Observed frequency	228	34	42	17	9	8	10

Table 10.1. Observed migration frequencies

10.3.2 Data description for `rochmigx.dta`

Number of observations (rows): 384
Number of level-2 cases: 384

10.3.3 Variables

`case`: case number
`n`: number of annual migrations since leaving school
`t`: number of years since leaving school
`ed`: educational qualification; a factor variable with 5 levels:

- 1=Degree or equivalent; professional qualifications with a degree
- 2=Education above A-level but below degree level; includes professional qualifications without a degree
- 3=A-level or equivalent
- 4=Other educational qualification
- 5=None

case	n	t	ed
50004	2	7	4
50008	0	3	4
50011	0	16	5
50016	5	9	4
50018	1	22	3
50020	0	21	5
50026	7	32	2
50028	0	31	4
50032	2	39	5
50046	0	5	4
50047	0	38	3
50057	0	12	5
50060	0	28	2
50064	0	2	3
50069	0	29	4
50071	0	8	2
50074	0	4	5
50075	0	11	5
50077	0	7	3
50079	7	26	3
50084	5	32	4

First few lines of `rochmigx.dta`

To model heterogeneity in migration propensity due to unmeasured and unmeasurable factors, we use a Poisson GLMM. To see if there is an inflated number of zeros in the count data, we allow for the left endpoint ($S_l = [0]$). In the Sabre `do` file below, you will notice that after transferring the data to Sabre, we reverse the coding of education, estimate a non-random effects model, a random effects model and a random effects model with the left endpoint.

Sabre commands

```
log using rochmigx_s.log, replace
set more off
use rochmigx
sabre, data case n t ed
sabre case n t ed, read
sabre, case case
sabre, yvar n
sabre, family p
sabre, trans logt log t
sabre, trans ned ed - 6
sabre, trans reved ned * -1
sabre, fac reved fed
sabre, constant cons
sabre, lfit cons logt
sabre, dis m
sabre, dis e
sabre, lfit cons logt fed
sabre, dis m
sabre, dis e
sabre, quad a
sabre, mass 24
sabre, fit cons logt
sabre, dis m
sabre, dis e
sabre, fit cons logt fed
sabre, dis m
sabre, dis e
sabre, end l
sabre, fit cons logt
sabre, dis m
sabre, dis e
sabre, fit cons logt fed
sabre, dis m
sabre, dis e
log close
clear
exit
```

Sabre log file

```
X-vars            Y-var             Case-var
------------------------------------------------
cons              n                 case
logt
fed

Univariate model
Standard Poisson
```

```
Gaussian random effects

Number of observations            =     348
Number of cases                   =     348

X-var df          =     6
Scale df          =     1

Log likelihood =     -417.15252     on     341 residual degrees of freedom


Parameter              Estimate        Std. Err.
___________________________________________________
cons                    -4.7394        0.62820
logt                     1.2670        0.18110
fed         ( 1)         0.0000        ALIASED [I]
fed         ( 2)        0.25333        0.24925
fed         ( 3)      -0.26454E-01     0.42727
fed         ( 4)        0.68149        0.42215
fed         ( 5)        0.67896        0.36583
scale                    1.2401        0.12924


Univariate model
Standard Poisson
Gaussian random effects, with left endpoint

Number of observations            =     348
Number of cases                   =     348

X-var df          =     6
Scale df          =     1
Endpoint df       =     1

Log likelihood =     -404.73727     on     340 residual degrees of freedom


Parameter              Estimate        Std. Err.
___________________________________________________
cons                    -2.6920        0.57807
logt                    0.97278        0.15614
fed         ( 1)         0.0000        ALIASED [I]
fed         ( 2)        0.44241        0.18511
fed         ( 3)      -0.33954E-01     0.32195
fed         ( 4)        0.67488        0.32357
fed         ( 5)        0.32689        0.27761
scale                   0.44976        0.13011
                                                        PROBABILITY
                                                        ___________
endpoint 0              0.92766        0.19008          0.48124
```

The log file shows that the random effects model with endpoints (stayers) has an improved log likelihood (-404.73727), when compared to the random effects model without stayers (-417.15252). In this case, the difference in log-likelihoods is not chi-square distributed, as under the null hypothesis, the

$\Pr[S_l] = [0]$ is on the edge of the parameter space. However, we can say that the probability that a randomly sampled individual is a stayer is estimated to be 0.48124.

The Poisson GLMM with an endpoint suggests: (1) that educational qualifications do significantly affect the likelihood of migration; (2) that there is evidence that the probability of migration varies markedly between individuals and (3) that the sample contains a highly significant number of "stayers".

With a single count of the number of annual migrations over an individual's working life, we can not distinguish between a heterogeneous population, with some individuals having a consistently high propensity to migrate and others a consistently low propensity to migrate, and a truly contagious process, i.e. one in which an individual's experience of migration increases the probability of subsequent migration.

The Poisson model assumes that the intervals between events are exponentially distributed, i.e. do not depend on duration of stay at a location. To examine this, we include duration in the next model.

10.3.4 Binary Response Model

In this part we use the data set `rochmig.dta` and model the individual binary response of whether or not there was a migration move in each calendar year.

10.3.5 Data description for `rochmig.dta`

Number of observations: 6349
Number of level-2 cases: 348

10.3.6 Variables

`case`: Case number
`move`: 1 if migration takes place in the year, 0 otherwise
`age`: age in years
`year`: calendar year
`dur`: duration of stay at each address

The data set (`rochmig.dta`) also contains a range of individual-specific covariates, though we do not use them in this particular exercise. These covariates include: education, employment status: esb2=1 (self employed), esb2=2 (employed), esb2=3 (not working); occupational status: osb3=1 (small proprietors, supervisors), osb3=0 (otherwise), promotion to service class: ops=0 (no), ops=1 (yes); first marriage: mfm=0 (no), mfm=1 (yes); marital break-up: mbu=0 (no), mbu=1 (yes); remarriage: mrm=0 (no), mrm=1 (yes); presence of children age 15-16: ch3=0 (no), ch3=1 (yes); marital status: msb2=0 (not married), msb2=1 (married).

case	move	age	year	dur	ed	ch1	ch2	ch3	ch4	msb	mse	esb	ese	osb	ose	mbu	mrm	mfm	msb1	epm	eoj	esb1	ops	osb1	msb2
50004	1	17	79	1	4	0	0	0	0	1	1	7	7	60	71	0	0	0	1	0	0	3	0	3	0
50004	0	18	80	1	4	0	0	0	0	1	1	7	7	71	71	0	0	0	1	0	0	3	0	2	0
50004	0	19	81	2	4	0	0	0	0	1	1	7	7	71	71	0	0	0	1	0	0	3	0	2	0
50004	1	20	82	3	4	0	0	0	0	1	1	7	7	71	60	0	0	0	1	0	0	3	0	2	0
50004	0	21	83	1	4	0	0	0	0	1	1	7	7	60	32	0	0	0	1	0	0	3	0	3	0
50004	0	22	84	2	4	0	0	0	0	1	1	7	0	32	0	0	0	0	1	0	0	3	0	6	0
50004	0	23	85	3	4	0	0	0	0	1	2	0	7	0	71	0	0	1	1	0	1	4	0	1	0
50008	0	18	83	1	4	0	0	0	0	1	1	7	0	31	0	0	0	0	1	0	0	3	0	6	0
50008	0	19	84	2	4	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	4	0	1	0
50008	0	20	85	3	4	0	0	0	0	1	1	0	7	0	71	0	0	0	1	0	1	4	0	1	0
50011	0	16	70	1	5	0	0	0	0	1	1	7	7	71	71	0	0	0	1	0	0	3	0	2	0
50011	0	17	71	2	5	0	0	0	0	1	1	7	7	71	71	0	0	0	1	0	0	3	0	2	0
50011	0	18	72	3	5	0	0	0	0	1	1	7	7	71	71	0	0	0	1	0	0	3	0	2	0
50011	0	19	73	4	5	0	0	0	0	1	1	7	7	71	71	0	0	0	1	0	0	3	0	2	0
50011	0	20	74	5	5	0	0	0	0	1	1	7	7	71	71	0	0	0	1	0	0	3	0	2	0
50011	0	21	75	6	5	0	0	0	0	1	1	7	7	71	71	0	0	0	1	0	0	3	0	2	0
50011	0	22	76	7	5	0	0	0	0	1	1	7	7	71	71	0	0	0	1	0	0	3	0	2	0
50011	0	23	77	8	5	0	0	0	0	1	1	7	7	71	71	0	0	0	1	0	0	3	0	2	0

First few lines of `rochmig.dta`

The sabre `do` file starts by transforming `age` and producing up to the 6th power of this transformed age effect (`stage`, stage2,..., `stage6`). The `do` file then estimates a range of homogeneous binary response models (logit link) before estimating a binary response GLMM using adaptive quadrature with 12 mass points. The final estimated model has lower and upper endpoints. The illustrated model was selected from a range of models estimated on these data, as can be seen from the TRAMSS web site,
http://tramss.data-archive.ac.uk/documentation/migration/migpag0.htm#Top.

Sabre commands

```
log using rochmig1_s.log, replace
set more off
use rochmig
#delimit ;
sabre, data case move age year dur ed ch1 ch2 ch3 ch4 msb mse esb ese osb
            ose mbu mrm mfm msb1 epm eoj esb1 ops osb1 msb2 esb2 osb2 osb3;
sabre case move age year dur ed ch1 ch2 ch3 ch4 msb mse esb ese osb ose mbu
      mrm mfm msb1 epm eoj esb1 ops osb1 msb2 esb2 osb2 osb3, read;
#delimit cr
sabre, case case
sabre, yvar move
sabre, fac dur fdur 5 10 15 20 25 30
sabre, fac year fyear 55 60 65 70 75 80
sabre, fac age fage 20 25 30 35 40 45
sabre, trans logdur log dur
sabre, trans year2 year * year
sabre, trans year3 year2 * year
sabre, trans age2 age * age
sabre, trans age3 age2 * age
sabre, trans age4 age3 * age
sabre, trans age5 age4 * age
sabre, trans age6 age5 * age
sabre, trans stage1 age - 30
sabre, trans stage stage1 / 10
sabre, trans stage2 stage * stage
sabre, trans stage3 stage2 * stage
sabre, trans stage4 stage3 * stage
```

```
sabre, trans stage5 stage4 * stage
sabre, trans stage6 stage5 * stage
sabre, constant cons
sabre, lfit cons fdur fyear fage
sabre, dis m
sabre, dis e
sabre, lfit cons dur year year2 year3 age age2 age3 age4 age5 age6
sabre, dis m
sabre, dis e
#delimit ;
sabre, lfit cons dur year year2 year3 stage stage2 stage3 stage4 stage5
            stage6;
#delimit cr
sabre, dis m
sabre, dis e
#delimit ;
sabre, lfit cons logdur year year2 year3 stage stage2 stage3 stage4 stage5
            stage6;
#delimit cr
sabre, dis m
sabre, dis e
sabre, lfit cons logdur year year2 stage stage2 stage3 stage4 stage5 stage6
sabre, dis m
sabre, dis e
sabre, lfit cons logdur year stage stage2 stage3 stage4 stage5 stage6
sabre, dis m
sabre, dis e
sabre, quad a
sabre, mass 12
sabre, fit cons logdur year stage stage2 stage3 stage4 stage5 stage6
sabre, dis m
sabre, dis e
sabre, end b
sabre, fit cons logdur year stage stage2 stage3 stage4 stage5 stage6
sabre, dis m
sabre, dis e
log close
clear
exit
```

Sabre log file

```
    Univariate model
    Standard logit
    Gaussian random effects

    Number of observations            =    6349
    Number of cases                   =     348

    X-var df          =     9
    Scale df          =     1

    Log likelihood =     -1072.0095     on    6339 residual degrees of freedom


    Parameter             Estimate        Std. Err.
    ___________________________________________________
    cons                    1.0442        0.77669
```

```
logdur              -0.69751         0.10795
year                -0.47716E-01     0.10365E-01
stage                0.63781E-01     0.34303
stage2               0.43262E-01     0.60000
stage3              -0.84894         0.53921
stage4               0.35929         0.55120
stage5               0.58352         0.21340
stage6              -0.28447         0.15054
scale                0.94736         0.16065


Univariate model
Standard logit
Gaussian random effects, with endpoints

Number of observations            =    6349
Number of cases                   =     348

X-var df           =     9
Scale df           =     1
Endpoint df        =     2

Log likelihood =     -1066.8979     on    6337 residual degrees of freedom


Parameter            Estimate        Std. Err.
___________________________________________________
cons                 0.83264         0.77005
logdur              -0.65921         0.10462
year                -0.36505E-01     0.10856E-01
stage               -0.69719E-01     0.34062
stage2               0.76791E-01     0.59487
stage3              -0.82208         0.53733
stage4               0.33145         0.54900
stage5               0.56759         0.21311
stage6              -0.27657         0.15032
scale                0.47666         0.17356
                                                      PROBABILITY
                                                      ___________
endpoint 0           0.56709         0.19694          0.36124
endpoint 1           0.27462E-02     0.46369E-02      0.17494E-02
```

By adding both endpoints to the binary response GLMM, the log-likelihood has increased from -1072.0095 to -1066.8979. The chi-square test is not strictly valid, as under the null hypothesis of no endpoints, the endpoint parameters lie on the edge of the parameter space. However, this change suggests that endpoints are needed. The probability of 0.36 associated with the left endpoint gives a measure of the proportion of "stayers" in the population, i.e. those individuals never likely to migrate. Examination of the parameter estimate and standard error of the right endpoint (and corresponding probability of 0.0017) suggests that this parameter (which estimates the proportion of the population migrating every year) could be set to zero.

The coefficient estimate of `logdur` (log duration) is negative, but is considerably smaller in magnitude than its effect in the simple logistic model. The coefficient of `logdur` measures cumulative inertia effects, and its value confirms that there is an increasing disinclination to move with increasing length of residence. Inference about duration effects can be misleading unless there is control for omitted variables (Heckman and Singer, 1984).

The random effects are significant in the binary response GLMM with endpoints: the scale parameter equals 0.47666 (s.e. 0.17356). We could improve our model of migration by adding explanatory variables which measure life cycle factors, such as marriage, occupation and employment status and the presence of children in the family. For this, and more details on the interpretation of the age effects in this model, see the TRAMSS site,
http://tramss.data-archive.ac.uk/documentation/migration/migpag0.htm#Top.

10.4 Exercises

There are three endpoint exercises to accompany this section. These exercises are: EP1 (binary response model of trade union membership); EP2 (Poisson model of fish catches) and EP3 (binary response model of female labour market participation).

10.5 References

Davies, R., and Crouchley, R., (1986), The Mover-Stayer Model Requiescat in Pace, Sociological Methods and Research, 14, 356-380

Goodman, L.A., (1961), Statistical methods for the mover stayer model, Journal of the American Statistical Association, 56, 841-868.

Greene, W., (1994), Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models, Stern School of Business Working Paper, EC-94-10.

Heckman, J.J., and Singer, B., (1984), A method for minimizing the impact of distributional assumptions in econometric models of duration data, Econometrica, 52, 271-320.

Lambert, D., (1992), Zero-inflated Poisson Regression, With an Application to Defects in Manufacturing, Technometrics 34, 1-14.

Singer, B., and Spillerman, S., (1976), Some methodological issues in the analysis of longitudinal surveys, Annals of Economic and Social Measurement, 5, 447-474.

Chapter 11

State Dependence Models

11.1 Introduction

Longitudinal and panel data on recurrent events are substantively important in social science research for two reasons. First, they provide some scope for extending control for variables that have been omitted from the analysis. For example, differencing provides a simple way of removing time-constant effects (both omitted and observed) from the analysis. Second, a distinctive feature of social science theory is that it postulates that behaviour and outcomes are typically influenced by previous behaviour and outcomes, that is, there is positive 'feedback' (e.g. the McGinnis (1968) 'axiom of cumulative inertia'). A frequently noted empirical regularity in the analysis of unemployment data is that those who were unemployed in the past (or have worked in the past) are more likely to be unemployed (or working) in the future (Heckman, 2001, p. 706). Heckman asks whether this is due to a causal effect of being unemployed (or working) or whether it is a manifestation of a stable trait. These two issues are related because inference about feedback effects are particularly prone to bias if the additional variation due to omitted variables (stable trait) is ignored. With dependence upon previous outcome, the explanatory variables representing the previous outcome will, for structural reasons, normally be correlated with omitted explanatory variables and therefore will always be subject to bias using conventional modelling methods. Understanding of this generic substantive issue dates back to the study of accident proneness by Bates and Neyman (1952) and has been discussed in many applied areas, including consumer behaviour (Massy et al., 1970) and voting behaviour (Davies and Crouchley, 1985).

An important attraction of longitudinal data is that, in principle, they make it possible to distinguish a key type of causality, namely state dependence {SD}, i.e. the dependence of current behaviour on earlier or related outcomes, from the confounding effects of unobserved heterogeneity {H}, or omitted variables and non-stationarity {NS}, i.e. changes in the scale and relative importance of the systematic relationships over time. Large sample sizes reduce the problems created by local maxima in disentangling the H, SD and NS effects.

Most observational schemes for collecting panel and other longitudinal data commence with the process already under way. They will therefore tend to have an informative start; the initial observed response is typically dependent upon pre-sample outcomes and unobserved variables. In contrast to time series analysis and, as explained by Anderson and Hsiao (1981), Heckman (1981a,b), Bhargava and Sargan (1983) and others, failure to allow for this informative start when SD and H are present will prejudice consistent parameter estimation. Various treatments of the initial conditions problem for recurrent events with SD using random effects for H have been proposed; see for example: Crouchley and Davies (2001), Wooldridge (2005), Alfo and Aitkin (2006), Kazemi and Crouchley (2006), Stewart (2007). We will concentrate on first order models for state dependence in linear, binary and count response sequences.

11.2 Motivational Example

The data in Table 11.1 were collected in a one-year panel study of depression and help-seeking behaviour in Los Angeles (Morgan et al, 1983). Adults were interviewed during the spring and summer of 1979 and re-interviewed at thee-monthly intervals. A respondent was classified as depressed if they scored >16 on a 20-item list of symptoms.

Season (i)				
y_{1j}	y_{2j}	y_{3j}	y_{4j}	Frequency
0	0	0	0	487
0	0	0	1	35
0	0	1	0	27
0	0	1	1	6
0	1	0	0	39
0	1	0	1	11
0	1	1	0	9
0	1	1	1	7
1	0	0	0	50
1	0	0	1	11
1	0	1	0	9
1	0	1	1	9
1	1	0	0	16
1	1	0	1	9
1	1	1	0	8
1	1	1	1	19

Table 11.1. Depression data from Morgan et al (1983)
Note: 1 = depressed, 0 = not depressed

Morgan et al (1983) concluded that there is strong temporal dependence in this binary depression measure and that the dependence is consistent with a mover-stayer process in which depression is a stationary, Bernoulli process for an 'at risk' subset of the population. Davies and Crouchley (1986) showed that a more general mixed Bernoulli model provides a significantly better fit to the data. However, by its very nature, depression is difficult to overcome suggesting that state dependence might explain at least some of the observed temporal dependence, although it remains an empirical issue whether true contagion extends over three months. We might also expect seasonal effects due to the weather.

In other words, what is the relative importance of state dependence (first order Markov), non-stationarity (seasonal effects) and unobserved heterogeneity (differences between the subjects) in the Morgan et al (1983) depression data?

In two-level GLMs, the subject-specific unobserved random effects u_{0j} are integrated out of the joint distribution for the responses to obtain the likelihood function. Thus

$$L\left(\gamma,\phi,\sigma_{u_0}^2|\mathbf{y},\mathbf{x},\mathbf{z}\right)=\prod_j\int\limits_{-\infty}^{+\infty}\prod_{i=1}^{T}g\left(y_{ij}\mid\theta_{ij},\phi\right)f\left(u_{0j}\mid\mathbf{x},\mathbf{z}\right)du_{0j},$$

where we have extended the notation of $f\left(u_{0j}\mid\mathbf{x},\mathbf{z}\right)$ to acknowledge the possibility that the multilevel random effects (u_{0j}) can depend on the regressors $(\mathbf{x},\mathbf{z})$. For notational simplicity, we have assumed that all the sequences are of the same length (T), though this can be easily relaxed by replacing T with T_j in the likelihood function.

To allow for state dependence (specifically first order Markov effects) we need to further augment our standard notation. We do this by adding the previous response (y_{i-1j}) to the linear predictor of the model for y_{ij} so that

$$\theta_{ij}=\gamma_{00}+\sum_{p=1}^{P}\gamma_{p0}x_{pij}+\sum_{q=1}^{Q}\gamma_{0q}z_{qj}+\delta y_{i-1j}+u_{0j},\quad i=2,...,T,$$

where δ is the new parameter associated with first order Markov state dependence. We also explicitly acknowledge this change to the GLM by writing the response model as $g\left(y_{ij}\mid y_{i-1j},\theta_{ij},\phi\right)$.
This treatment of state dependence can be appropriate for modelling ongoing responses. However it begs the question: what do we do about the first observation? In panel data, the data window usually samples an ongoing process and the information collected on the initial observation rarely contains all of the pre-sample response sequence and its determinants back to inception. The implications of this will be explored. For the moment, we will write the response model for the initial observed response y_{1j} as $g\left(y_{1j}\mid\theta_{1j},\phi^1\right)$ to allow the parameters and multilevel random effects for the initial response to be different to those of subsequent responses, so that

$$L\left(\gamma^1,\gamma,\delta,\phi^1,\phi,\sigma_{u_0}^2|\mathbf{y},\mathbf{x},\mathbf{z}\right)=$$

$$\prod_j\int\limits_{-\infty}^{+\infty}g\left(y_{1j}\mid\theta_{1j},\phi^1\right)\prod_{i=2}^{T}g\left(y_{ij}\mid y_{i-1j},\theta_{ij},\phi\right)f\left(u_{0j}\mid\mathbf{x},\mathbf{z}\right)du_{0j}.$$

To the responses

$$\mathbf{y}_j=\left[y_{1j,}y_{2j},y_{3j},...,y_T\right],$$

we can relate time-varying regressors

$$\mathbf{x}_j=\left[\mathbf{x}_{1j},\mathbf{x}_{2j,...,}\mathbf{x}_{Tj}\right],$$

and time-constant regressors

$$\mathbf{z}_j = [\mathbf{z}_j] .$$

In particular, for the initial response

$$\theta_{1j} = \gamma_{00}^1 + \sum_{p=1}^{P} \gamma_{p0}^1 x_{p1j} + \sum_{q=1}^{Q} \gamma_{0q}^1 z_{qj} + u_{0j}.$$

In this likelihood, we have the same random effect (u_{0j}) for both the initial response and subsequent responses. This assumption will be relaxed later.

If we omit the first term on the right hand side of this likelihood function, we have conditioning on the initial response. The data window interrupts an ongoing process, whereby the initial observation y_{1j} will, in part, be determined by u_{0j}, and this simplification may induce inferential error.

This problem was examined by Anderson and Hsiao (1981) for the linear model. They compared Ordinary Least Squares, Generalised Least Squares, and Maximum Likelihood Estimation (MLE) for a number of different cases. MLE has desirable asymptotic properties when time T or sample size N (or both) $\rightarrow \infty$. In conventional panel studies, T is fixed and often small. For random (i.e. endogenous) y_{1j}, only MLE provides consistent parameter estimation but this requires the inclusion of

$$g\left(y_{1j} \mid \theta_{1j}, \phi^1\right)$$

in the likelihood. Specification of this density is itself problematic for non-linear models, as emphasised by Diggle et al (1994, p193). Heckman (1981b) suggests using an approximate formulation including whatever covariates are available. Various treatments of the initial conditions problem for recurrent events with state dependence using random effects for heterogeneity have been proposed; see for example: Crouchley and Davies (2001), Wooldridge (2005), Alfo and Aitkin (2006), Kazemi and Crouchley (2006), Stewart (2007).

We will review the alternative treatments of the initial conditions problem and illustrate them on the binary depression data.

11.3 First Order Markov State Dependence in GLMs

11.3.1 Conditional Model: Conditional on the Initial Response

If we omit the model for the initial response from the likelihood, we get

$$L^c\left(\gamma, \delta, \phi, \sigma_{u_0}^2 | \mathbf{y}, x, \mathbf{z}\right) = \prod_j \int_{-\infty}^{+\infty} \prod_{i=2}^{T} g\left(y_{ij} \mid y_{i-1j}, \theta_{ij}, \phi\right) f\left(u_{0j} \mid \mathbf{x}, \mathbf{z}\right) du_{0j}.$$

For the responses

$$\mathbf{y}_j = [y_{2j}, y_{3j}, ..., y_{Tj}] ,$$

we have included the lagged response in the time-varying regressors

$$\mathbf{x}_{ij} = [x_{ij}, y_{i-1j}],$$

$$\mathbf{z}_j = [z_j].$$

Further we are ignoring any dependence of the random effects on the regressors:

$$f(u_{0j} \mid \mathbf{x}, \mathbf{z}) = f(u_{0j}).$$

The above likelihood simplifies to

$$L^c\left(\gamma, \delta, \phi, \sigma^2_{u_0} | \mathbf{y}, \mathbf{x}, \mathbf{z}\right) = \prod_j \int_{-\infty}^{+\infty} \prod_{i=2}^{T} g\left(y_{ij} \mid y_{i-1j}, \theta_{ij}, \phi\right) f\left(u_{0j}\right) du_{0j},$$

with

$$\theta_{ij} = \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + \delta y_{i-1j} + u_{0j}$$

for $i = 2, ..., T$.

11.4 Depression example

In this section, we will estimate a model for the binary depression data `depression2.dta`. This is a version of the data without the initial response. The model includes a constant, dummy variables for seasons 3 and 4 and the lagged response variable.

11.4.1 Data description for `depression2.dta`

Number of observations: 2256
Number of level-2 cases: 752

11.4.2 Variables

`ind`: individual indentifier
`t`: season
`t2`: 1 if t=2, 0 otherwise
`t3`: 1 if t=3, 0 otherwise
`t4`: 1 if t=4, 0 otherwise
`s`: 1 if the respondent is depressed, 0 otherwise
`s1`: baseline response
`s_lag1`:lag 1 response
`s_lag2`:lag 2 response (not used)

ind	t	t2	t3	t4	s	s1	s_lag1	s_lag2
1	2	1	0	0	0	0	0	-9
1	3	0	1	0	0	0	0	0
1	4	0	0	1	0	0	0	0
2	2	1	0	0	0	0	0	-9
2	3	0	1	0	0	0	0	0
2	4	0	0	1	0	0	0	0
3	2	1	0	0	0	0	0	-9
3	3	0	1	0	0	0	0	0
3	4	0	0	1	0	0	0	0
4	2	1	0	0	0	0	0	-9
4	3	0	1	0	0	0	0	0
4	4	0	0	1	0	0	0	0
5	2	1	0	0	0	0	0	-9
5	3	0	1	0	0	0	0	0
5	4	0	0	1	0	0	0	0
6	2	1	0	0	0	0	0	-9
6	3	0	1	0	0	0	0	0

Figure 11.1: First few lines of `depression2.dta`

11.4.3 Sabre commands

```
log using depression1_s.log, replace
set more off
use depression2
sabre, data ind t t2 t3 t4 s s1 s_lag1 s_lag2
sabre ind t t2 t3 t4 s s1 s_lag1 s_lag2, read
sabre, case ind
sabre, yvar s
sabre, link p
sabre, constant cons
sabre, lfit t3 t4 s_lag1 cons
sabre, dis m
sabre, dis e
sabre, mass 24
sabre, fit t3 t4 s_lag1 cons
sabre, dis m
sabre, dis e
log close
clear
exit
```

11.4.4 Sabre log file

```
The first few lines of depression2.dta


(Standard Homogenous Model)
```

```
    Parameter              Estimate        Std. Err.
    ___________________________________________________
    (intercept)           -1.2636          0.61874E-01
    t3                    -0.13649         0.84561E-01
    t4                    -0.15150E-02     0.82817E-01
    s_lag1                 1.0480          0.79436E-01


(Random Effects Model)


    Parameter              Estimate        Std. Err.
    ___________________________________________________
    (intercept)           -1.2942          0.72379E-01
    t3                    -0.15466         0.88638E-01
    t4                    -0.21480E-01     0.87270E-01
    s_lag1                 0.94558         0.13563
    scale                  0.32841         0.18226

    X-vars            Y-var            Case-var
    ________________________________________________
    (intercept)       response         case.1
    t3
    t4
    s_lag1

    Univariate model
    Standard probit
    Gaussian random effects

    Number of observations            =    2256
    Number of cases                   =     752

    X-var df          =     4
    Scale df          =     1

    Log likelihood =     -831.56731     on    2251 residual degrees of freedom
```

11.4.5 Discussion

The coefficient on y_{i-1j} (s_lag1) is 0.94558 (s.e. 0.13563), which is highly significant, but the scale parameter (σ) is of marginal significance, suggesting a nearly homogeneous first order model. Can we trust this inference?

11.5 Conditioning on the initial response but allowing the random effect u_{0j} to be dependent on $\mathbf{z}_j$, Wooldridge (2005)

Wooldridge (2005) proposes that we drop the term $g\left(y_{1j} \mid \theta_{1j}, \phi^1\right)$ and use the conditional likelihood

$$L^c\left(\gamma, \delta, \phi, \sigma_{u_0}^2 | \mathbf{y}, \mathbf{x}, \mathbf{z}\right) = \prod_j \int_{-\infty}^{+\infty} \prod_{i=2}^{T} g\left(y_{ij} \mid y_{i-1j}, \theta_{ij}, \phi\right) f\left(u_{0j} \mid \mathbf{x}, \mathbf{z}\right) du_{0j},$$

where

$$\mathbf{y}_j = \left[y_{2j}, y_{3j}, ..., y_{Tj}\right],$$
$$\mathbf{z}_j = \left[z_j, y_{1j}\right],$$
$$\mathbf{x}_{ij} = \left[x_{ij}, y_{i-1j}\right],$$

but rather than assume u_{0j} is iid, i.e. $f\left(u_{0j} \mid \mathbf{x}, \mathbf{z}\right) = f\left(u_{0j}\right)$ as in Section 11.3.1 we use

$$f\left(u_{0j} \mid \mathbf{x}, \mathbf{z}\right) = f\left(u_{0j} \mid \mathbf{z}\right).$$

By allowing the omitted (random) effects to depend on the initial response

$$u_{0j} = \kappa_{00} + \kappa_1 y_{1j} + \sum_{q=1}^{Q} \kappa_{0q} z_{qj} + u_{0j}^{\mathrm{w}},$$

where u_{0j}^{w} is independent and identically distributed, we get

$$\begin{aligned}
\theta_{ij} &= \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + \delta y_{i-1j} + u_{0j} \\
&= \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + \delta y_{i-1j} + \kappa_{00} + \kappa_1 y_{1j} + \sum_{q=1}^{Q} \kappa_{0q} z_{qj} + u_{0j}^{\mathrm{w}} \\
&= (\gamma_{00} + \kappa_{00}) + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} (\gamma_{0q} + \kappa_{0q}) z_{qj} + \delta y_{i-1j} + \kappa_1 y_{1j} + u_{0j}^{\mathrm{w}} \\
&= \gamma_{00}^{\mathrm{w}} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \gamma_{0q}^{\mathrm{w}} z_{qj} + \delta y_{i-1j} + \kappa_1 y_{1j} + u_{0j}^{\mathrm{w}}.
\end{aligned}$$

This implies that coefficients on the constant $(\gamma_{00} + \kappa_{00}) = \gamma_{00}^{\mathrm{w}}$ and the time-constant covariates $(\gamma_{0q} + \kappa_{0q}) = \gamma_{0q}^{\mathrm{w}}$ will be confounded. The ability of the auxiliary model

$$u_{0j} = \kappa_{00} + \kappa_1 y_{1j} + \sum_{q=1}^{Q} \kappa_{0q} z_{qj} + u_{0j}^{\mathrm{w}}$$

to account for the dependence in $f(u_{0j} \mid \mathbf{x}, \mathbf{z})$ will depend to some extent on the nature of the response (y_{ij}). For binary initial responses (y_{1j}) only one parameter κ_1 is needed, but for the linear model and count data, polynomials in y_{1j} may be needed to account more fully for the dependence. Also, as Wooldridge (2005) suggests, we can include interaction effects between the y_{1j} and z_{qj}.

Crouchley and Davies (1999) raise inferential issues about the inclusion of baseline responses (initial conditions) in models without state dependence.

11.6 Depression example

The model for the binary depression data `depression2.dta`, ignoring the model for the initial response, has constant, dummy variables for seasons 3 and 4, the lagged response variable and the initial response.

11.6.1 Sabre commands

```
log using depression2_s.log, replace
set more off
use depression2
sabre, data ind t t2 t3 t4 s s1 s_lag1 s_lag2
sabre ind t t2 t3 t4 s s1 s_lag1 s_lag2, read
sabre, case ind
sabre, yvar s
sabre, link p
sabre, constant cons
sabre, lfit t3 t4 s_lag1 s1 cons
sabre, dis m
sabre, dis e
sabre, mass 24
sabre, fit t3 t4 s_lag1 s1 cons
sabre, dis m
sabre, dis e
log close
clear
exit
```

11.6.2 Sabre log file

```
(Standard Homogenous Model)

    Parameter             Estimate        Std. Err.
    ___________________________________________________
    (intercept)          -1.3390          0.65010E-01
```

```
t3                    -0.12914        0.85893E-01
t4                    -0.70059E-02    0.84373E-01
s_lag1                 0.69132        0.96958E-01
s1                     0.62535        0.93226E-01
```

```
(Random Effects Model)


    Parameter              Estimate         Std. Err.
    ___________________________________________________
    (intercept)           -1.6646          0.11654
    t3                    -0.20988         0.99663E-01
    t4                    -0.88079E-01     0.97569E-01
    s_lag1                 0.43759E-01     0.15898
    s1                     1.2873          0.19087
    scale                  0.88018         0.12553

    X-vars            Y-var             Case-var
    ________________________________________________
    (intercept)       response          case.1
    t3
    t4
    s_lag1
    s1

    Univariate model
    Standard probit
    Gaussian random effects

    Number of observations             =    2256
    Number of cases                    =     752

    X-var df           =     5
    Scale df           =     1

    Log likelihood =     -794.75310     on    2250 residual degrees of freedom
```

11.6.3 Discussion

This model has the lagged response s_lag1 estimate at 0.043759 (s.e. 0.15898), which is not significant, while the initial response s1 estimate 1.2873 (s.e. 0.19087) and the scale parameter estimate 0.88018 (s.e. 0.12553) are highly significant. There is also a big improvement in the log-likelihood over the model without `s1` of

$$-2(-831.56731 - (-794.75310)) = 73.628$$

for 1 degree of freedom. This model has no time-constant covariates to be confounded by the auxiliary model and suggests that depression is a zero-order process.

11.7 Modelling the initial conditions

There are several approximations that can be adopted: (1) use the same random effect in the initial and subsequent responses, e.g. Crouchley and Davies (2001); (2) use a one-factor decomposition for the initial and subsequent responses, e.g. Heckman (1981a), Stewart (2007); (3) use different (but correlated) random effects for the initial and subsequent responses; (4) embed the Wooldridge (2005) approach in joint models for the initial and subsequent responses.

Same random effect in the initial and subsequent responses with a common scale parameter

The likelihood for this model is

$$L\left(\gamma^1, \gamma, \delta, \phi^1, \phi, \sigma_{u0}^2 | \mathbf{y}, \mathbf{x}, \mathbf{z}\right) =$$

$$\prod_j \int_{-\infty}^{+\infty} g\left(y_{1j} \mid \theta_{1j}, \phi^1\right) \prod_{i=2}^{T} g\left(y_{ij} \mid y_{i-1j}, \theta_{ij}, \phi\right) f\left(u_{0j} \mid \mathbf{x}, \mathbf{z}\right) du_{0j},$$

where the responses

$$\mathbf{y}_j = \left[y_{1j}, y_{2j}, y_{3j}, ..., y_{Tj}\right],$$

time-varying regressors

$$\mathbf{x}_j = \left[\mathbf{x}_{1j}, \mathbf{x}_{2,j,...,}\mathbf{x}_{Tj}\right],$$

time-constant regressors

$$\mathbf{z}_j = \left[z_j\right].$$

For the initial response

$$\theta_{1j} = \gamma_{00}^1 + \sum_{p=1}^{P} \gamma_{p0}^1 x_{p1j} + \sum_{q=1}^{Q} \gamma_{0q}^1 z_{qj} + u_{0j},$$

and for subsequent responses we have

$$\theta_{ij} = \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + \delta y_{i-1j} + u_{0j}, i = 2, ..., T.$$

To set this model up in Sabre, we combine the linear predictors by using dummy variables so that for all i

$$\begin{aligned}
\theta_{ij} &= r_1 \theta_{1j} + r_2 \theta_{ij},\ i = 2, ..., T, \\
r_1 &= 1, \text{if } i = 1, 0 \text{ otherwise}, \\
r_2 &= 1, \text{if } i > 1, 0 \text{ otherwise},
\end{aligned}$$

where for all i

$$\text{var}\left(u_{0j}\right) = \sigma_{u0}^2.$$

For the binary and Poisson models, we have $\phi = 1$ in $g\left(y_{ij} \mid y_{i-1j}, \theta_{ij}, \phi\right)$, for the linear model, we have

$$\phi = \sigma^2_{\varepsilon^1}$$

for the initial response, and

$$\phi = \sigma^2_{\varepsilon}$$

for subsequent responses.

11.7.1 Depression example

The joint model for the binary depression data `depression.dta` has a constant for the initial response, a constant for the subsequent responses, dummy variables for seasons 3 and 4 and the lagged response variable.

11.7.2 Data description for `depression.dta`

Number of observations: 3008
Number of level-2 cases: 752

11.7.3 Variables

`ind`: individual identifier
`t`: season (1,2,3,4)
`t1`: 1 if t=1, 0 otherwise
`t2`: 1 if t=2, 0 otherwise
`t3`: 1 if t=3, 0 otherwise
`t4`: 1 if t=4, 0 otherwise
`s`: 1 if the respondent is depressed, 0 otherwise
`s1`: baseline response
`s_lag1`: lag 1 response, -9 if missing
`s_lag2`: lag 2 response, -9 if missing (not used)
`r`: response position, 1 if baseline, 2 if subsequent response
`r1`: 1 if r=1, 0 otherwise
`r2`: 1 if r=2, 0 otherwise

In the depression example, the model for the initial response (indicator `r1`=1), has only a constant, the model for the 3 subsequent responses (indicator `r2`=1) has a constant, dummy variables for seasons 3 (`r2_t3`) and 4 (`r2_t4`), and the lagged response variable (`r2_lag1`).

11.7.4 Sabre commands

```
log using depression3_s.log, replace
```

ind	t	t1	t2	t3	t4	s	s1	s_lag1	s_lag2	r	r1	r2
1	1	1	0	0	0	0	0	-9	-9	1	1	0
1	2	0	1	0	0	0	0	0	-9	2	0	1
1	3	0	0	1	0	0	0	0	0	2	0	1
1	4	0	0	0	1	0	0	0	0	2	0	1
2	1	1	0	0	0	0	0	-9	-9	1	1	0
2	2	0	1	0	0	0	0	0	-9	2	0	1
2	3	0	0	1	0	0	0	0	0	2	0	1
2	4	0	0	0	1	0	0	0	0	2	0	1
3	1	1	0	0	0	0	0	-9	-9	1	1	0
3	2	0	1	0	0	0	0	0	-9	2	0	1
3	3	0	0	1	0	0	0	0	0	2	0	1
3	4	0	0	0	1	0	0	0	0	2	0	1
4	1	1	0	0	0	0	0	-9	-9	1	1	0
4	2	0	1	0	0	0	0	0	-9	2	0	1
4	3	0	0	1	0	0	0	0	0	2	0	1
4	4	0	0	0	1	0	0	0	0	2	0	1
5	1	1	0	0	0	0	0	-9	-9	1	1	0
5	2	0	1	0	0	0	0	0	-9	2	0	1
5	3	0	0	1	0	0	0	0	0	2	0	1
5	4	0	0	0	1	0	0	0	0	2	0	1
6	1	1	0	0	0	0	0	-9	-9	1	1	0
6	2	0	1	0	0	0	0	0	-9	2	0	1

Figure 11.2: First few lines of `depression.dta`

```
set more off
use depression
sabre, data ind t t1 t2 t3 t4 s s1 s_lag1 s_lag2 r r1 r2
sabre ind t t1 t2 t3 t4 s s1 s_lag1 s_lag2 r r1 r2, read
sabre, case ind
sabre, yvar s
sabre, link p
sabre, trans r2_t3 r2 * t3
sabre, trans r2_t4 r2 * t4
sabre, trans r2_lag1 r2 * s_lag1
sabre, lfit r1 r2 r2_t3 r2_t4 r2_lag1
sabre, dis m
sabre, dis e
sabre, mass 24
sabre, fit r1 r2 r2_t3 r2_t4 r2_lag1
sabre, dis m
sabre, dis e
log close
clear
exit
```

11.7.5 Sabre log file

```
(Standard Homogenous Model)


    Parameter              Estimate         Std. Err.
    ___________________________________________________
    r1                    -0.93769          0.53811E-01
    r2                    -1.2636           0.61874E-01
    r2_t3                 -0.13649          0.84561E-01
    r2_t4                 -0.15150E-02      0.82817E-01
    r2_lag1                1.0480           0.79436E-01


(Random Effects Model)


    Parameter              Estimate         Std. Err.
    ___________________________________________________
    r1                    -1.3476           0.10026
    r2                    -1.4708           0.92548E-01
    r2_t3                 -0.20740          0.99001E-01
    r2_t4                 -0.85438E-01      0.97129E-01
    r2_lag1                0.70228E-01      0.14048
    scale                  1.0372           0.10552

    X-vars            Y-var             Case-var
    ________________________________________________
    r1                response          case.1
    r2
    r2_t3
    r2_t4
    r2_lag1
```

```
Univariate model
Standard probit
Gaussian random effects

Number of observations            =    3008
Number of cases                   =     752

X-var df          =     5
Scale df          =     1

Log likelihood =     -1142.9749     on    3002 residual degrees of freedom
```

11.7.6 Discussion

The non-significant coefficient of `r2_lag1` 0.070228 (s.e. 0.14048) suggests that there is no state dependence in these data, while the highly significant scale coefficient 1.0372 (s.e. 0.10552) suggests heterogeneity.

11.8 Same random effect in models of the initial and subsequent responses but with different scale parameters

This model can be derived from a one-factor decomposition of the random effects for the initial and subsequent observations; for its use in this context, see Heckman (1981a) and Stewart (2007). The likelihood for this model

$$L\left(\gamma^1, \gamma, \delta, \phi^1, \phi, \sigma^2_{1u0}, \sigma^2_{u_0} | \mathbf{y}, \mathbf{x}, \mathbf{z}\right),$$

is just like that for the common scale parameter model with the same random effect for the initial and subsequent responses except that for $i = 1$, we have

$$\text{var}\left(u_{0j}\right) = \sigma^2_{1u0}$$

and for $i > 1$,

$$\text{var}\left(u_{0j}\right) = \sigma^2_{u0}.$$

In binary or linear models, the scale parameter for the initial response is identified from the covariance of y_{1j} and the y_{ij}, $i > 1$. Stewart (2007) has a different parameterization for $i = 1$:

$$\text{var}\left(u_{0j}\right) = \lambda\sigma^2_{u0}$$

and for $i > 1$,

$$\text{var}\left(u_{0j}\right) = \sigma^2_{u0}.$$

As in the common scale parameter model we combine the linear predictors by using dummy variables so that for all i

$$\theta_{ij} = r_1\theta_{1j} + r_2\theta_{ij},\ i = 2, ..., T,$$
$$r_1 = 1, \text{if } i = 1, 0 \text{ otherwise},$$
$$r_2 = 1, \text{if } i > 1, 0 \text{ otherwise}.$$

11.9 Depression example

As in the common scale parameter model, the joint model for the binary depression data `depression.dta` has a constant for the initial response, a constant for the subsequent responses, dummy variables for seasons 3 and 4 and the lagged response variable.

11.9.1 Sabre commands

```
log using depression4_s.log, replace
set more off
use depression
sabre, data ind t t1 t2 t3 t4 s s1 s_lag1 s_lag2 r r1 r2
sabre ind t t1 t2 t3 t4 s s1 s_lag1 s_lag2 r r1 r2, read
sabre, case ind
sabre, yvar s
sabre, rvar r
sabre, link p
sabre, trans r2_t3 r2 * t3
sabre, trans r2_t4 r2 * t4
sabre, trans r2_lag1 r2 * s_lag1
sabre, nvar 1
sabre, lfit r1 r2 r2_t3 r2_t4 r2_lag1
sabre, dis m
sabre, dis e
sabre, mass first=24 second=24
sabre, depend y
sabre, nvar 1
sabre, fit r1 r2 r2_t3 r2_t4 r2_lag1
sabre, dis m
sabre, dis e
log close
clear
exit
```

11.9.2 Sabre log file

```
    Univariate model
    Standard probit

    Number of observations             =    3008
```

```
X-var df          =     5

Log likelihood =     -1179.8909     on    3003 residual degrees of freedom


Parameter              Estimate         Std. Err.
___________________________________________________
r1                    -0.93769          0.53811E-01
r2                    -1.2636           0.61874E-01
r2_t3                 -0.13649          0.84561E-01
r2_t4                 -0.15150E-02      0.82817E-01
r2_lag1                1.0480           0.79436E-01



Univariate model
Dependent probit
Gaussian random effects

Number of observations            =    3008
Number of cases                   =     752

X-var df          =     5
Scale df          =     2

Log likelihood =     -1142.9355     on    3001 residual degrees of freedom


Parameter              Estimate         Std. Err.
___________________________________________________
r1                    -1.3248           0.12492
r2                    -1.4846           0.10639
r2_t3                 -0.21020          0.10004
r2_t4                 -0.87882E-01      0.98018E-01
r2_lag1                0.50254E-01      0.15792
scale1                 1.0021           0.15927
scale2                 1.0652           0.14587
```

11.9.3 Discussion

This shows that the state dependence regressor `r2_lag1` has estimate 0.050254 (s.e. 0.15792), which is not significant. It also shows that the scale parameters $\left(\sigma^2_{1u0}, \sigma^2_{u0}\right)$ are nearly the same. The log-likelihood improvement of the model with 2 scale parameters over that of the previous model with one scale parameter is

$$-2(-1142.9749 - (-1142.9355)) = 0.0788,$$

for 1 degree of freedom Thus the model with 1 scale parameter is to be preferred.

11.10 Different random effects in models of the initial and subsequent responses

The likelihood for this model is

$$L\left(\gamma^1,\gamma,\delta,\phi^1,\phi,\sigma^2_{u_0},\rho|\mathbf{y},\mathbf{x},\mathbf{z}\right)=$$

$$\prod_j \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g\left(y_{1j}\mid\theta_{1j},\phi^1\right)\prod_{i=2}^{I} g\left(y_{ij}\mid y_{i-1j},\theta_{ij},\phi\right) f\left(u^1_{0j},u^2_{0j}\mid\mathbf{x},\mathbf{z}\right)du^1_{0j}du^2_{0j},$$

where the responses

$$\mathbf{y}_j=\left[y_{1j,}y_{2j},y_{3j},...,y_{Tj}\right],$$

time-varying regressors

$$\mathbf{x}_j=\left[\mathbf{x}_{1j},\mathbf{x}_{2j,...,}\mathbf{x}_{Tj}\right],$$

time-constant regressors

$$\mathbf{z}_j=\left[\mathbf{z}_j\right].$$

The main difference between this joint model and the previous single random effect version is the use of different random effects for the initial and subsequent responses. This implies that we need a bivariate integral to form the marginal likelihood. For the initial response

$$\theta_{1j}=\gamma^1_{00}+\sum_{p=1}^{P}\gamma^1_{p0}x_{p1j}+\sum_{q=1}^{Q}\gamma^1_{0q}z_{qj}+u^1_{0j},$$

and for subsequent responses we have

$$\theta_{ij}=\gamma_{00}+\sum_{p=1}^{P}\gamma_{p0}x_{pij}+\sum_{q=1}^{Q}\gamma_{0q}z_{qj}+\delta y_{i-1j}+u^2_{0j},,i=2,...,T.$$

The correlation between the random effects $\left(u^1_{0j},u^2_{0j}\right)$ has parameter ρ, which is identified from the covariance of y_{1j} and the y_{ij}, $i>1$. The scale parameter for the initial response is not identified in the presence of ρ in the binary or linear models, so in these models we hold it at the same value as that of the subsequent responses.

As in all joint models, to set this model up in Sabre, we combine the linear predictors by using dummy variables so that for all i

$$\begin{aligned}\theta_{ij}&=r_1\theta_{1j}+r_2\theta_{ij},\ i=2,...,T,\\ r_1&=1,\text{if } i=1,0\text{ otherwise},\\ r_2&=1,\text{if } i>1,0\text{ otherwise}.\end{aligned}$$

For the binary and Poisson models, we have $\phi=1$ in $g\left(y_{ij}\mid y_{i-1j},\theta_{ij},\phi\right)$. For the linear model, we still have

$$\phi=\sigma^2_{\varepsilon^1}$$

for the initial response, and

$$\phi = \sigma_{\varepsilon}^2$$

for subsequent responses.

11.11 Depression example

As in the single random effect models, the joint model for the binary depression data `depression.dta` has a constant for the initial response, a constant for the subsequent responses, dummy variables for seasons 3 and 4 and the lagged response variable.

11.11.1 Sabre commands

```
log using depression5_s.log, replace
set more off
use depression
sabre, data ind t t1 t2 t3 t4 s s1 s_lag1 s_lag2 r r1 r2
sabre ind t t1 t2 t3 t4 s s1 s_lag1 s_lag2 r r1 r2, read
sabre, case ind
sabre, yvar s
sabre, model b
sabre, rvar r
sabre, link first=p second=p
sabre, trans r2_t3 r2 * t3
sabre, trans r2_t4 r2 * t4
sabre, trans r2_lag1 r2 * s_lag1
sabre, nvar 1
sabre, lfit r1 r2 r2_t3 r2_t4 r2_lag1
sabre, dis m
sabre, dis e
sabre, mass 24
sabre, eqscale y
sabre, der1 y
sabre, nvar 1
sabre, fit r1 r2 r2_t3 r2_t4 r2_lag1
sabre, dis m
sabre, dis e
log close
clear
exit
```

11.11.2 Sabre log file

```
Standard probit/probit

Number of observations             =    3008

X-var df          =     5
```

```
Log likelihood =     -1179.8909     on    3003 residual degrees of freedom


Parameter              Estimate         Std. Err.
___________________________________________________
r1                    -0.93769         0.53811E-01
r2                    -1.2636          0.61874E-01
r2_t3                 -0.13649         0.84561E-01
r2_t4                 -0.15150E-02     0.82817E-01
r2_lag1                1.0480          0.79436E-01


Correlated bivariate model

Standard probit/probit
Gaussian random effects

Number of observations             =    3008
Number of cases                    =     752

X-var df           =     5
Scale df           =     2

Log likelihood =     -1142.9355     on    3001 residual degrees of freedom


Parameter              Estimate         Std. Err.
___________________________________________________
r1                    -1.3672          0.12386
r2                    -1.4846          0.10591
r2_t3                 -0.21020         0.10033
r2_t4                 -0.87881E-01     0.97890E-01
r2_lag1                0.50253E-01     0.15946
scale                  1.0652          0.14362
corr                   0.97091         0.10087
```

11.11.3 Discussion

Note that the log-likelihood is exactly the same as that for the previous model. The `scale2` parameter from the previous model has the same value as the scale parameter of the current model. The lagged response `r2_lag1` has an estimate of 0.050313 (s.e. 0.15945), which is not significant. The correlation between the random effects (`corr`) has estimate 0.97089 (s.e. 0.10093), which is very close to 1 suggesting that the common random effects, zero-order, single scale parameter model is to be preferred.

11.12 Embedding the Wooldridge (2005) approach in joint models for the initial and subsequent responses

This extended model will help us to assess the value of the Wooldridge (2005) approach in an empirical context. We can include the initial response in the linear predictors of the subsequent responses of any of the joint models, but for simplicity we will use the single random effect, single scale parameter model.

The likelihood for this model is

$$L\left(\gamma^1, \gamma, \delta, \phi^1, \phi, \sigma^2_{u_0} | \mathbf{y}, \mathbf{x}, \mathbf{z}^{\mathrm{P}}\right) =$$

$$\prod_j \int_{-\infty}^{+\infty} g\left(y_{1j} \mid \theta_{1j}, \phi^1\right) \prod_{i=2}^{T} g\left(y_{ij} \mid y_{i-1j}, y_{1j}, \theta_{ij}, \phi\right) f\left(u_{0j} \mid \mathbf{x}, \mathbf{z}^{\mathrm{P}}\right) du_{0j},$$

where the responses

$$\mathbf{y}_j = \left[y_{1j}, y_{2j}, y_{3j}, ..., y_{Tj}\right],$$

time-varying regressors

$$\mathbf{x}_j = \left[\mathbf{x}_{1j}, \mathbf{x}_{2j,...,} \mathbf{x}_{Tj}\right],$$

time-constant regressors

$$\mathbf{z}_j = \left[z_j, y_{1j}\right].$$

For the initial response

$$\theta_{1j} = \gamma^1_{00} + \sum_{p=1}^{P} \gamma^1_{p0} x_{p1j} + \sum_{q=1}^{Q} \gamma^1_{0q} z_{qj} + u_{0j},$$

and for subsequent responses we have

$$\theta_{ij} = \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + \delta y_{i-1j} + \kappa_1 y_{1j} + u_{0j}, i = 2, ..., T,$$

as we have added $\kappa_1 y_{1j}$ to the linear predictor.

As with joint models, we combine the linear predictors by using dummy variables so that for all i

$$\theta_{ij} = r_1 \theta_{1j} + r_2 \theta_{ij}, \ i = 2, ..., T,$$
$$r_1 = 1, \text{if } i = 1, 0 \text{ otherwise},$$
$$r_2 = 1, \text{if } i > 1, 0 \text{ otherwise},$$

where for all i

$$\mathrm{var}\left(u_{0j}\right) = \sigma^2_{u0}.$$

For the binary and Poisson models, we have $\phi = 1$ in $g\left(y_{ij} \mid y_{i-1j}, \theta_{ij}, \phi\right)$, for the linear model, we have

$$\phi = \sigma^2_{\varepsilon^1}$$

for the initial response, and

$$\phi = \sigma_\varepsilon^2$$

for subsequent responses.

11.13 Depression example

As in the single random effect models, the joint model for the binary depression data `depression.dta` has a constant for the initial response, a constant for the subsequent responses, dummy variables for seasons 3 and 4, the lagged response variable and the initial response variable.

11.13.1 Sabre commands

```
log using depression6_s.log, replace
set more off
use depression
sabre, data ind t t1 t2 t3 t4 s s1 s_lag1 s_lag2 r r1 r2
sabre ind t t1 t2 t3 t4 s s1 s_lag1 s_lag2 r r1 r2, read
sabre, case ind
sabre, yvar s
sabre, link p
sabre, trans r2_t3 r2 * t3
sabre, trans r2_t4 r2 * t4
sabre, trans r2_lag1 r2 * s_lag1
sabre, trans r2_base r2 * s1
sabre, lfit r1 r2 r2_t3 r2_t4 r2_lag1 r2_base
sabre, dis m
sabre, dis e
sabre, mass 24
sabre, fit r1 r2 r2_t3 r2_t4 r2_lag1 r2_base
sabre, dis m
sabre, dis e
log close
clear
exit
```

11.13.2 Sabre log file

```
(Standard Homogenous Model)

    Parameter              Estimate         Std. Err.
    ___________________________________________________
    r1                    -0.93769         0.53811E-01
    r2                    -1.3390          0.65010E-01
    r2_t3                 -0.12914         0.85893E-01
    r2_t4                 -0.70059E-02     0.84373E-01
```

```
r2_lag1              0.69132         0.96958E-01
r2_s1                0.62535         0.93226E-01


(Random Effects Model)

    Parameter             Estimate        Std. Err.
    ___________________________________________________
    r1                   -1.3632          0.16189
    r2                   -1.4741          0.97129E-01
    r2_t3                -0.20869         0.99797E-01
    r2_t4                -0.86541E-01     0.97774E-01
    r2_lag1               0.61491E-01     0.15683
    r2_s1                -0.33542E-01     0.26899
    scale                 1.0602          0.21274

    X-vars           Y-var            Case-var
    ________________________________________________
    r1               response         case.1
    r2
    r2_t3
    r2_t4
    r2_lag1
    r2_s1

    Univariate model
    Standard probit
    Gaussian random effects

    Number of observations            =    3008
    Number of cases                   =     752

    X-var df          =     6
    Scale df          =     1

    Log likelihood =     -1142.9670     on    3001 residual degrees of freedom
```

11.13.3 Discussion

This joint model has both the lagged response `r2_lag1` estimate of 0.061490 (s.e. 0.15683) and the baseline/initial response effect `r2_base` estimate of -0.033544 (s.e. 0.26899) as being non-significant.

11.14 Other link functions

State dependence can also occur in Poisson and linear models. For linear model examples, see Baltagi and Levin (1992) and Baltagi (2005). These data concern the demand for cigarettes per capita by state for 46 American States.

We have found first-order state dependence in the Poisson data of Hall et al. (1984), Hall, Griliches and Hausman (1986). The data refer to the number of patents awarded to 346 firms each year from 1975 to 1979.

11.15 Exercises

There are a range of exercises to accompany this chapter. The exercises FOL1, FOL2, FOL3 and FOC2 are for binary responses. The exercises FOC4 and FOC5 are for Poisson responses. These exercises show that the Wooldridge (2005) approach works well for binary responses, but (in its simplest form) not for Poisson data.

11.16 References

Alfò M., & Aitkin, M., (2006), Variance component models for longitudinal count data with baseline information: epilepsy data revisited. Statistics and Computing, Volume 16, 231-238

Anderson, T.W., & Hsiao, C., (1981), Estimation of dynamic models with error components, JASA, 76, 598-606.

Baltagi, B.H., Levin, D. (1992), "Cigarette taxation: raising revenues and reducing consumption", Structural Change and Economic Dynamics, Vol. 3 pp.321-35

Baltagi, Badi H. (2005) Econometric Analysis of Panel Data, West Sussex: John Wileyand Sons.

Bates, G.E., & Neyman, J., (1952), Contributions to the theory of accident proneness, I, An optimistic model of the correlation between light & severe accidents, II, True or false contagion, Univ Calif, Pub Stat, 26, 705-720.

Bhargava, A. & Sargan, J.D., (1983), Estimating dynamic random effects models from panel data covering short time periods, Econometrica, 51, 1635-1657.

Crouchley, R., & Davies, R.B., (2001), A comparison of GEE & random effects models for distinguishing hetero-geneity, nonstationarity & state dependence in a collection of short binary event series, Stat Mod, 1, 271-285.

Crouchley, R., & Davies, R.B., (1999), A comparison of population average and random effects models for the analysis of longitudinal count data with baseline information, JRSS, A, 162, 331-347.

Davies, R.B., Elias, P., and Penn, R., (1992), The relationship between a husband's unemployment and his wife's participation in the labour force, Oxford

Bulletin of Economics and Statistics, 54, 145-171

Davies, R.B., (1993), Statistical modelling for survey analysis, Journal of the Market Research Society, 35, 235-247.

Davies, R.B. & Crouchley, R., (1985), The determinants of party loyalty: a disaggregate analysis of panel data from the 1974 and 1979 General Elections in England, Political Geography Quarterly, 4, 307-320.

Davies, R.B. & Crouchley, R., (1986), The mover-stayer model: Requiescat in pace, *Sociological Methods and Research*, 14, 356-380.

Diggle, P.J., Liang, K. Y. & Zeger, S. L., (1994), *Anaysis of Longitudinal data*, Clarendon Press, Oxford.

Hausman, J., Hall, B., & Z. Griliches, Z., (1984), Econometric models for count data with an application to the patents – R&D relationship, Econometrica, 52, 909-938.

Hall,B, Zvi Griliches, and Jerry Hausman (1986), "Patents and R&D: Is There a Lag?", International Economic Review, 27, 265-283.

Heckman J.J., (1981a), Statistical models for discrete panel data, In Manski, C.F. & McFadden, D, (eds), Structural analysis of discrete data with econometric applications, MIT press, Cambridge, Mass.

Heckman J.J., (1981b), The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process, In Manski, C.F. & McFadden, D, (eds), Structural analysis of discrete data with econometric applications, MIT press, Cambridge, Mass.

Heckman J.J., (2001), "Micro data, heterogeneity and the evaluation of public policy: Nobel lecture", Journal of Political Economy, 109, 673—748.

Kazemi, I., & Crouchley, R., (2006), Modelling the initial conditions in dynamic regression models of panel data with random effects, Ch 4, in Baltagi, B.H., Panel Data Econometrics, theoretical Contributions and Empirical Applications, Elsevier, Amsterdam, Netherlands.

Massy, W.F., Montgomery, D.B., & Morrison, D.G., (1970), Stochastic models of buying behaviour, MIT Press, Cambridge, Mass.

McGinnis, R., (1968), A stochastic model of social mobility, American Sociological Review, 23, 712-722.

Morgan, T.M., Aneshensel, C.S. & Clark, V.A. (1983), Parameter estimation for mover stayer models: analysis of depression over time, *Soc Methods and Research*, 11, 345-366.

Rabe-Hesketh, S., & Skrondal, A., (2005), Multilevel and Longitudinal Modelling using Stata, Stata Press, Stata Corp, College Station, Texas

Stewart, M.B., (2007), The interrelated dynamics of unemployment and low-wage employment, Journal of Applied Econometrics, Volume 22, 511-531

Vella, F.,& Verbeek, M., (1998), Whose wages do Unions raise? A dynamic Model of Unionism and wage rate determination for young men, Journal of Applied Econometrics, 13, 163-183

Wooldridge, J.M., (2005), Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity, Journal of Applied Econometrics, 20, 39—54.

Chapter 12

Incidental Parameters: An Empirical Comparison of Fixed Effects and Random Effects Models

12.1 Introduction

The main objective of the random effects/multilevel modelling approach is the estimation of the γ parameters in the presence of the random effects or incidental parameters (in a 2-level model these are the individual specific random effects u_{0j}). This has been done by assuming that the incidental parameters are Gaussian distributed, and by computing the expected behaviour of individuals randomly sampled from this distribution (in other words, by integrating the random effects out of the model). For the 2-level random effects generalised linear mixed model we had the likelihood

$$L\left(\gamma, \phi, \sigma_{u_0}^2 | \mathbf{y}, \mathbf{x}, \mathbf{z}\right) = \prod_j \int \prod_i g\left(y_{ij} \mid \theta_{ij}, \phi\right) f\left(u_{0j}\right) du_{0j},$$

where

$$g\left(y_{ij} \mid \theta_{ij}, \phi\right) = \exp\left\{\left[y_{ij}\theta_{ij} - b\left(\theta_{ij}\right)\right]/\phi + c\left(y_{ij}, \phi\right)\right\},$$

$$\theta_{ij} = \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + u_{0j},$$

and

$$f\left(u_{0j}\right) = \frac{1}{\sqrt{2\pi}\sigma_{u_0}} \exp\left(-\frac{u_{0j}^2}{2\sigma_{u_0}^2}\right).$$

This approach will provide consistent estimates of the $\gamma = [\gamma_{00}, \gamma_{p0}, \gamma_{0q}]$ so long as in the true model, the u_{0j} are independent of the covariates $[x, z]$.

A second approach, due to Andersen (1973), is to find a sufficient statistic for the u_{0j} and to estimate the γ from a likelihood conditional upon this sufficient statistic. For the binary response model with a logit link, the formulation uses the probability of the grouped responses conditional upon $S_j = \sum_i y_{ij}$ (for panel data, this is the total number or count of events observed for that individual over the observation period). The distribution of the data $y_{1j}, ..., y_{Tj}$ conditional on S_j is free of u_{0j}. The product of these conditional distributions provides a likelihood whose maximum will provide a consistent estimator of γ_{p0}. The γ_{00} and γ_{0q} are conditioned out of the likelihood. The same approach can be used with the Poisson model. When there is some form of state dependence or endogeneity in the binary response model, the conditional likelihood approach gives inconsistent estimates, see Crouchley and Pickles (1988).

There are several other related approaches. One involves factoring the likelihood into two orthogonal parts, one for the structural parameters and another for the incidental parameters, e.g. Cox and Reid (1987). Another related approach is to estimate the u_{0j} and some of the elements of γ by the usual maximum likelihood procedures. For instance, with panel/clustered data, only the parameters on the time/within cluster varying covariates (γ_{p0}) in the linear model are identified.

In a panel, the number of panel members is large and the period of observation is short. As the number of panel members increases, so too does the number of incidental parameters (u_{0j}). This feature was called the "incidental parameters problem" by Neyman and Scott (1948). For the linear model, with only time/within cluster varying covariates, maximum likelihood gives consistent γ_{p0} but biased u_{0j}.

Abowd et al (2002) developed an algorithm for the the direct least squares estimation of (γ_{p0}, u_{0j}) in linear models on very large data sets. `FEFIT` is the Sabre version of this algorithm. The estimates of u_{0j} produced by direct least squares are consistent as the cluster size or $T_j \to \infty$, see Hsiao (1986, section 3.2) and Wooldridge (2002, Ch10). The number of dummy variables (u_{0j}) that can be directly estimated using conventional matrix manipulation in least squares is limited by storage requirements, so `FEFIT` uses sparse matrix procedures. `FEFIT` has been tested on some very large data sets (e.g. with over 1 million fixed effects). `FEFIT` has been written in a way that enables it to use multiple processors in parallel.

We start by reviewing the fixed effects (FE) treatment of the 2-level linear model and show how to estimate this model in Sabre, we then compare the FE with the random effects (RE) model. The Chapter ends with a discussion about the 3-level FE model.

12.2 Fixed Effects Treatment of The 2-Level Linear Model

Using the notation of Chapter 3, the explanatory variables at the individual level (level 1) are denoted by $x_1, \cdots, x_P$, and those at the group level (level 2) by $z_1, \cdots, z_Q$. This leads to the following formula

$$y_{ij} = \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + u_{0j} + \varepsilon_{ij},$$

where the regression parameters γ_{p0} $(p = 1, \cdots, P)$ and γ_{0q} $(q = 1, \cdots, Q)$ are for level-one and level-two explanatory variables, respectively. If we treat the incidental parameters u_{0j} as fixed effects or constants, then without additional restrictions, the γ_{0q}, γ_{00}, and u_{0j} are not separately identifiable or estimable.

If we absorb the z_{qj} into the fixed effect, so that

$$y_{ij} = \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + u_{0j}^{+} + \varepsilon_{ij},$$

where

$$u_{0j}^{+} = \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + u_{0j}.$$

Then we can identify the fixed effects u_{0j}^{+} by introducing the restriction $\sum u_{0j}^{+} = 0$. The individual incidental parameter u_{0j}^{+} represents the deviation of the jth individual from the common mean γ_{00}. Another way to identify the u_{0j}^{+} is to treat them as dummy variables and set one to zero, i.e. put it in the reference group (alternatively drop the constant from the model). The fixed effect u_{0j}^{+} may be correlated with the included explanatory variables x_{pij} (unlike the random effect version). We still assume that the residuals ε_{ij} are mutually independent and have zero means conditional on the explanatory variables. The population variance of the level-one residuals ε_{ij} is denoted by σ_{ε}^2.

We can form a mean version (over i) of the model for y_{ij} so that

$$\overline{y}_j = \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} \overline{x}_{pj} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + u_{0j} + \overline{\varepsilon}_j,$$

where $\overline{x}_{pj} = \sum x_{pij}/T_j$, $\overline{y}_j = \sum y_{ij}/T_j$, $\overline{\varepsilon}_j = \sum \varepsilon_{ij}$, $z_{qj} = \sum z_{qj}/T_j$ and $u_{0j} = \sum u_{0j}/T_j$. The mean version (over i) of the model still contains the original constant, cluster or time constant covariates and the incidental parameter u_{0j}. The mean version (over i) of the model for y_{ij}, produces one observation for each individual or cluster. The u_{0j} are not identified in this model as they occur only once in each cluster and are absorbed into the residual.

If we take the mean model from the basic form we get what is called the demeaned model with clustered data or time-demeaned model with longitudinal data, i.e.

$$\left(y_{ij}-\overline{y}_j\right)=\sum_{p=1}^{P}\gamma_{p0}\left(x_{pij}-\overline{x}_{pj}\right)+\left(\varepsilon_{ij}-\overline{\varepsilon}_j\right).$$

This differenced form does not have a constant, any incidental parameters, or group-specific (time-constant) covariates in its specification. This differenced or demeaned form is often estimated using OLS.

The random effects and fixed effects models can lead to quite different inference about γ_{p0}. For example, Hausman (1978) found that using a fixed effects estimator produced significantly different results from a random effects specification of a wage equation. Mundlak (1978a) suggested that, in the random effects formulation, we approximate $E\left(u_{0j}\mid \mathbf{x}_{pj}\right)$ by a linear function, i.e.

$$u_{0j}=\sum_{p=1}^{P}\gamma_{p0}^{*}\overline{x}_{pj}+\omega_{oj},$$

where $\omega_{oj}\sim N(0,\sigma_{\omega}^{2})$, see also Chamberlain (1980). So that

$$y_{ij}=\gamma_{00}+\sum_{p=1}^{P}\gamma_{p0}^{*}\overline{x}_{pj}+\sum_{p=1}^{P}\gamma_{p0}x_{pij}+\sum_{q=1}^{Q}\gamma_{0q}z_{qj}+\omega_{0j}+\varepsilon_{ij}.$$

Mundlak (1978) suggests that if we use this augmented GLMM, then the difference between the random and fixed effects specifications would disappear. However, there is another explanation of why there could be differences between the two formulations. Suppose we had the alternative augmented GLMM,

$$y_{ij}=\gamma_{00}+\sum_{p=1}^{P}\gamma_{p0}^{**}\overline{x}_{pj}+\sum_{p=1}^{P}\gamma_{p0}^{+}\left(x_{pij}-\overline{x}_{pj}\right)+\sum_{q=1}^{Q}\gamma_{0q}z_{qj}+u_{0j}+\varepsilon_{ij},$$

which reduces to the original form if $\gamma_{p0}^{**}=\gamma_{p0}^{+}$. In this model, a change in the average value of $\overline{x}_{pj}$ has a different impact to differences from the average. The mean form (over i) of the alternative augmented model gives

$$\overline{y}_j=\gamma_{00}+\sum_{p=1}^{P}\gamma_{p0}^{**}\overline{x}_{pj}+\sum_{q=1}^{Q}\gamma_{0q}z_{qj}+u_{0j}+\overline{\varepsilon}_j.$$

If we take this mean form (over i) from the alternative augmented model, then we get the time-demeaned form,

$$\left(y_{ij}-\overline{y}_j\right)=\sum_{p=1}^{P}\gamma_{p0}^{+}\left(x_{pij}-\overline{x}_{pj}\right)+\left(\varepsilon_{ij}-\overline{\varepsilon}_j\right).$$

In this case, the mean and time-demeaned forms are estimating different things.

Hausman and Taylor (1981) show how to identify time-varying effects using a fixed effects estimator and identify the time-constant effects using a random effects estimator in the same regression. This specification is currently beyond the scope of Sabre.

12.2.1 Dummy Variable Specification of the Fixed Effects Model

Hsiao (1986, section 3.2) shows that by using dummy variables for the incidental parameters in a linear model with time-varying covariates, i.e.

$$y_{ij} = \sum_{p=1}^{P} \gamma_{p0} x_{pij} + u_{0j}^{*} + \varepsilon_{ij},$$

we can obtain the same estimates as those of the differenced model

$$\left(y_{ij} - \overline{y}_{j}\right) = \sum_{p=1}^{P} \gamma_{p0} \left(x_{pij} - \overline{x}_{pj}\right) + \left(\varepsilon_{ij} - \overline{\varepsilon}_{j}\right).$$

However, the differenced model parameter estimates will have smaller standard errors, unless the calculation of the means $\left(\overline{y}_{j}, \overline{x}_{pj}\right)$ is taken into account. The OLS estimates of the fixed effects are given by

$$\widehat{u}_{0j}^{*} = \overline{y}_{j} - \sum_{p=1}^{P} \gamma_{p0} \overline{x}_{pj}.$$

Sabre has a procedure `FEFIT` which uses least squares to directly estimate the dummy variable version of the incidental parameter model. One advantage of the dummy variable form of the model is that it can be applied to the non demeaned data when the level-2 nesting is broken, e.g. when pupils (level 1) change class (level 2).

12.3 Empirical Comparison of 2-Level Fixed and Random Effects Estimators

We now empirically compare the various ways of estimating a linear model with incidental parameters. The data we use are a version of the National Longitudinal Study of Youth (NLSY) as used in various Stata Manuals (to illustrate the `xt` commands). The data are for young women who were aged 14-26 in 1968. The women were surveyed each year from 1970 to 1988, except for 1974, 1976, 1979, 1981, 1984 and 1986. We have removed records with missing values on the response (log wages) and explanatory variables. There are 4132 women (`idcode`) with between 1 and 12 years of data on being in waged employment (i.e. not in full-time education) and earning over \$1/hour and less than \$700/hour. We are going to explore how the results change when we use different estimators of the incidental parameters.

12.3.1 References

Stata Longitudinal/Panel Data, Reference Manual, Release 9, (2005), Stata Press, StataCorp LP, College Station, Texas.

12.3.2 Data description for `nlswork.dta`

Number of observations: 28091
Number of level-2 cases: 4132

12.3.3 Variables

`ln_wage`: ln(wage/GNP deflator) in a particular year
`black`: 1 if woman is black, 0 otherwise
`msp`: 1 if woman is married and spouse is present, 0 otherwise
`grade`: years of schooling completed (0-18)
`not_smsa`: 1 if woman was living outside a standard metropolitan statistical area (smsa), 0 otherwise
`south`: 1 if woman was living in the South, 0 otherwise
`union`: 1 if woman was a member of a trade union, 0 otherwise
`tenure`: job tenure in years (0-26)
`age`: respondent's age
`age2`: age* age

idcode	year	birth_yr	age	race	msp	nev_mar	grade	collgrad	not_smsa	c_city	south	union	ttl_exp	tenure	ln_wage	black	age2	ttl_exp2	tenure2
1	72	51	20	2	1	0	12	0	0	1	0	1	2.26	0.92	1.59	1	400	5.09	0.84
1	77	51	25	2	0	0	12	0	0	1	0	0	3.78	1.50	1.78	1	625	14.26	2.25
1	80	51	28	2	0	0	12	0	0	1	0	1	5.29	1.83	2.55	1	784	28.04	3.36
1	83	51	31	2	0	0	12	0	0	1	0	1	5.29	0.67	2.42	1	961	28.04	0.44
1	85	51	33	2	0	0	12	0	0	1	0	1	7.16	1.92	2.61	1	1089	51.27	3.67
1	87	51	35	2	0	0	12	0	0	0	0	1	8.99	3.92	2.54	1	1225	80.77	15.34
1	88	51	37	2	0	0	12	0	0	0	0	1	10.33	5.33	2.46	1	1369	106.78	28.44
2	71	51	19	2	1	0	12	0	0	1	0	0	0.71	0.25	1.36	1	361	0.51	0.06
2	77	51	25	2	1	0	12	0	0	1	0	1	3.21	2.67	1.73	1	625	10.31	7.11
2	78	51	26	2	1	0	12	0	0	1	0	1	4.21	3.67	1.69	1	676	17.74	13.44
2	80	51	28	2	1	0	12	0	0	1	0	1	6.10	5.58	1.73	1	784	37.16	31.17
2	82	51	30	2	1	0	12	0	0	1	0	1	7.67	7.67	1.81	1	900	58.78	58.78
2	83	51	31	2	1	0	12	0	0	1	0	1	8.58	8.58	1.86	1	961	73.67	73.67
2	85	51	33	2	0	0	12	0	0	1	0	1	10.18	1.83	1.79	1	1089	103.62	3.36
2	87	51	35	2	0	0	12	0	0	1	0	1	12.18	3.75	1.85	1	1225	148.34	14.06
2	88	51	37	2	0	0	12	0	0	1	0	1	13.62	5.25	1.86	1	1369	185.55	27.56
3	71	45	25	2	0	1	12	0	0	1	0	0	3.44	1.42	1.55	1	625	11.85	2.01
3	72	45	26	2	0	1	12	0	0	1	0	0	4.44	2.42	1.61	1	676	19.73	5.84
3	73	45	27	2	0	1	12	0	0	1	0	0	5.38	3.33	1.60	1	729	28.99	11.11
3	77	45	31	2	0	1	12	0	0	1	0	0	6.94	2.42	1.62	1	961	48.20	5.84

First few lines of `nlswork.dta`

The version of the data set that we use has the time-demeaned covariates (denoted *var*`tilde,` e.g. agetilde) included.

Sabre Commands: homogeneous model

These commands open a log file, read the data and estimate a homogeneous linear model with time-varying covariates and finally close the log file.

```
log using nlswork_fe_nolfit_s.log, replace
set mem 100m
set more off
use nlswork
#delimit ;
sabre, data idcode year birth_yr age race grade collgrad not_smsa c_city
            south ttl_exp tenure ln_wage age2 ttl_exp2 tenure2 black
            ln_wagebar ln_wagetilde gradebar gradetilde agebar agetilde
            age2bar age2tilde ttl_expbar ttl_exptilde ttl_exp2bar
            ttl_exp2tilde tenurebar tenuretilde tenure2bar tenure2tilde
            blackbar blacktilde not_smsabar not_smsatilde southbar
            southtilde;
sabre idcode year birth_yr age race grade collgrad not_smsa c_city south
      ttl_exp tenure ln_wage age2 ttl_exp2 tenure2 black ln_wagebar
      ln_wagetilde gradebar gradetilde agebar agetilde age2bar age2tilde
      ttl_expbar ttl_exptilde ttl_exp2bar ttl_exp2tilde tenurebar
      tenuretilde tenure2bar tenure2tilde blackbar blacktilde not_smsabar
      not_smsatilde southbar southtilde,read;
#delimit cr
sabre, case idcode
sabre, yvar ln_wage
sabre, fam g
sabre, constant cons
#delimit ;
sabre, lfit age age2 ttl_exp ttl_exp2 tenure tenure2 not_smsa south grade
            black cons;
#delimit cr
sabre, dis m
sabre, dis e
log close
clear
exit
```

Sabre Log File: Homogeneous Linear Model

```
    Univariate model
    Standard linear
    Number of observations             =   28091

    X-var df           =    11
    Sigma df           =     1

    Log likelihood =     -12523.347     on   28079 residual degrees of freedom


    Parameter              Estimate         Std. Err.
    ___________________________________________________
    cons                   0.24728          0.49332E-01
    age                    0.38598E-01      0.34670E-02
    age2                  -0.70818E-03      0.56322E-04
    ttl_exp                0.21128E-01      0.23350E-02
```

```
ttl_exp2             0.44733E-03      0.12461E-03
tenure               0.47369E-01      0.19626E-02
tenure2             -0.20270E-02      0.13380E-03
not_smsa            -0.17205          0.51675E-02
south               -0.10034          0.48938E-02
grade                0.62924E-01      0.10313E-02
black               -0.69939E-01      0.53207E-02
sigma                0.37797
```

Sabre Commands: Time-Demeaned Data and Model

These are the extra commands needed to estimate a homogeneous linear model with time-demeaned covariates.

```
sabre, yvar ln_wagetilde
#delimit ;
sabre, lfit agetilde age2tilde ttl_exptilde ttl_exp2tilde tenuretilde
            tenure2tilde not_smsatilde southtilde gradetilde blacktilde;
#delimit cr
sabre, dis m
sabre, dis e
```

Sabre Log File: demeaned model

```
Univariate model
Standard linear

Number of observations             =   28091

X-var df           =     8
Sigma df           =     1

Log likelihood =     -2578.2531     on   28082 residual degrees of freedom

Parameter              Estimate         Std. Err.
___________________________________________________
agetilde                0.35999E-01      0.30903E-02
age2tilde              -0.72299E-03      0.48601E-04
ttl_exptilde            0.33467E-01      0.27060E-02
ttl_exp2tilde           0.21627E-03      0.11657E-03
tenuretilde             0.35754E-01      0.16870E-02
tenure2tilde           -0.19701E-02      0.11406E-03
not_smsatilde          -0.89011E-01      0.86980E-02
southtilde             -0.60631E-01      0.99759E-02
gradetilde              0.0000           ALIASED [E]
blacktilde              0.0000           ALIASED [E]
sigma                   0.26527
```

Sabre Commands: Explicit Dummy Variables Model

These are the extra commands needed to estimate a homogeneous linear model with explicit dummy variables for the incidental individual-specific parameters (`fidcode`).

```
sabre, YVAR ln_wage
sabre, fam g
sabre, fac idcode fidcode
sabre, LFIT age age2 ttl_exp ttl_exp2 tenure tenure2 not_smsa south fidcode
```

Sabre Log File: Explicit Dummy Variables Model

```
Number of observations            =   28091

X-var df          =  4705
Sigma df          =     1

Log likelihood =     -2578.2532     on   23385 residual degrees of freedom


Parameter              Estimate        Std. Err.
___________________________________________________
age                    0.35999E-01      0.33864E-02
age2                  -0.72299E-03      0.53258E-04
ttl_exp                0.33467E-01      0.29653E-02
ttl_exp2               0.21627E-03      0.12774E-03
tenure                 0.35754E-01      0.18487E-02
tenure2               -0.19701E-02      0.12499E-03
not_smsa              -0.89011E-01      0.95316E-02
south                 -0.60631E-01      0.10932E-01
fidcode     (   1)        1.4233          0.96326E-01
fidcode     (   2)       0.97264          0.96648E-01
fidcode     (   3)       0.82992          0.89323E-01
fidcode     (   4)        1.3009          0.10013
fidcode     (   5)        1.1761          0.10011
fidcode     (   6)        1.0522          0.91844E-01
etc.
```

Discussion 1

As can be seen from the above log files, the model for the time-demeaned data and the explicit dummy variable model with the non-time-demeaned data produce identical estimates. These are both slightly different to those from the homogeneous model. If the incidental parameters are independent of the covariates, both sets of estimates will tend to the same limit as the number of clusters increases.

The covariates `gradetilde` and `blacktilde` are dropped from the time-demeaned model as these are time-constant covariates, which when demeaned have the value zero throughout. The smaller standard errors of the demeaned model parameter estimates occur because the model-fitting procedure has not taken into account the separate estimation of the means that were used to obtain the time-demeaned values.

12.3.4 Implicit Fixed Effects Estimator

This procedure (FEFIT) uses dummy variables for each individual, and solves the least squares normal equations using sparse matrix procedures. We call this the implicit fixed effects estimator, as the dummy variables are not written out as part of the display.

Sabre Commands: Implicit Fixed Effects Model

These are the extra commands needed to estimate and display the results for the implicit fixed effects model.

```
sabre, yvar ln_wage
sabre, fefit age age2 ttl_exp ttl_exp2 tenure tenure2 not_smsa south
sabre, dis m
sabre, dis e
```

Sabre Log Files: Implicit Fixed Effects Model

```
Univariate model
Standard linear
Fixed effects

Number of observations            =   28091
Number of cases                   =    4697

X-var df          =     8

Parameter              Estimate        Std. Err.
___________________________________________________
age                    0.35999E-01     0.33865E-02
age2                  -0.72299E-03     0.53259E-04
ttl_exp                0.33467E-01     0.29654E-02
ttl_exp2               0.21627E-03     0.12774E-03
tenure                 0.35754E-01     0.18487E-02
tenure2               -0.19701E-02     0.12499E-03
not_smsa              -0.89011E-01     0.95318E-02
south                 -0.60631E-01     0.10932E-01
sigma                  0.29070
```

Discussion 2

This implicit dummy variable model does not have a constant. The estimates and standard errors match those of the explicit dummy variables model. Clearly with small data sets like the `nlswage.dta,` both the implicit and explicit dummy variable models can be used. However, the implicit model estimator `FEFIT` was 3000 times faster on this data set than `LFIT` and required much less memory.

12.3.5 Random Effects Models

We now use Sabre to obtain the RE estimates for the various specifications.

The Classical Random Effects Model

$$y_{ij} = \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + u_{0j} + \varepsilon_{ij}.$$

Sabre Commands: Classical Random Effects Model. These are the extra commands needed to estimate the model with $x_{pij} + z_{qj}$, using 6-point adaptive quadrature.

```
sabre, quad a
sabre, mass 6
#delimit ;
sabre, fit age age2 ttl_exp ttl_exp2 tenure tenure2 not_smsa south grade
        black cons;
#delimit cr
sabre, dis m
sabre, dis e
```

Sabre Log File

```
    Univariate model
    Standard linear
    Gaussian random effects

    Number of observations             =   28091
    Number of cases                    =    4697

    X-var df          =    11
    Sigma df          =     1
    Scale df          =     1

    Log likelihood =     -8853.4259     on   28078 residual degrees of freedom


    Parameter              Estimate         Std. Err.
```

```
 ___________________________________________________
 cons                      0.23908          0.49190E-01
 age                       0.36853E-01      0.31226E-02
 age2                     -0.71316E-03      0.50070E-04
 ttl_exp                   0.28820E-01      0.24143E-02
 ttl_exp2                  0.30899E-03      0.11630E-03
 tenure                    0.39437E-01      0.17604E-02
 tenure2                  -0.20052E-02      0.11955E-03
 not_smsa                 -0.13234          0.71322E-02
 south                    -0.87560E-01      0.72143E-02
 grade                     0.64609E-01      0.17372E-02
 black                    -0.53339E-01      0.97338E-02
 sigma                     0.29185          0.13520E-02
 scale                     0.24856          0.35017E-02
```

The Extended Random Effects Model 1

In this extension both the time means of the covariates $(\overline{x}_{pj})$ and the time-varying covariates (x_{pij}), have their own parameters in the linear predictor, i.e.

$$y_{ij} = \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0}^{*} \overline{x}_{pj} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + \omega_{0j} + \varepsilon_{ij}.$$

Sabre Commands: Extended Random Effects Model 1. These are the extra commands needed to estimate the model with $\overline{x}_{pj} + x_{pij} + z_{qj}$.

```
sabre, quad a
sabre, mass 6
#delimit ;
sabre, fit agebar age2bar ttl_expbar ttl_exp2bar tenurebar tenure2bar
           not_smsabar southbar age age2 ttl_exp ttl_exp2 tenure tenure2
           not_smsa south grade black cons;
#delimit cr
sabre, dis m
sabre, dis e
```

Sabre Log File

```
    Univariate model
    Standard linear
    Gaussian random effects

    Number of observations             =   28091
    Number of cases                    =    4697

    X-var df           =    19
    Sigma df           =     1
    Scale df           =     1

    Log likelihood =     -8774.6178     on   28070 residual degrees of freedom
```

```
Parameter              Estimate        Std. Err.
___________________________________________________
cons                    0.31033        0.12438
agebar                 -0.20870E-02    0.95809E-02
age2bar                 0.10329E-03    0.15613E-03
ttl_expbar             -0.19474E-01    0.63847E-02
ttl_exp2bar             0.49153E-03    0.34887E-03
tenurebar               0.31656E-01    0.62217E-02
tenure2bar             -0.79062E-03    0.42178E-03
not_smsabar            -0.98306E-01    0.14231E-01
southbar               -0.40645E-01    0.14537E-01
age                     0.35999E-01    0.33967E-02
age2                   -0.72299E-03    0.53421E-04
ttl_exp                 0.33467E-01    0.29744E-02
ttl_exp2                0.21627E-03    0.12813E-03
tenure                  0.35754E-01    0.18543E-02
tenure2                -0.19701E-02    0.12537E-03
not_smsa               -0.89011E-01    0.95607E-02
south                  -0.60631E-01    0.10965E-01
grade                   0.61112E-01    0.19098E-02
black                  -0.60684E-01    0.98738E-02
sigma                   0.29158        0.13489E-02
scale                   0.24458        0.34461E-02
```

The Extended Random Effects Model 2

In this extension both the time means of the covariates ($\overline{x}_{pj}$) and the time-demeaned covariates ($x_{pij} - \overline{x}_{pj}$) have their own parameters in the linear predictor, i.e.

$$y_{ij} = \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0}^{**} \overline{x}_{pj} + \sum_{p=1}^{P} \gamma_{p0}^{+} \left(x_{pij} - \overline{x}_{pj} \right) + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + u_{0j} + \varepsilon_{ij}.$$

Sabre Commands: Extended Random Effects Model 2. These are the extra commands needed to estimate the model with $\overline{x}_{pj} + (x_{pij} - \overline{x}_{pj}) + z_{qj}$.

```
sabre, quad a
sabre, mass 6
#delimit ;
sabre, fit agebar age2bar ttl_expbar ttl_exp2bar tenurebar tenure2bar
           not_smsabar southbar agetilde age2tilde ttl_exptilde
           ttl_exp2tilde tenuretilde tenure2tilde not_smsatilde southtilde
           grade black cons;
#delimit cr
sabre, dis m
sabre, dis e
```

Sabre Log File

```
Univariate model
Standard linear
Gaussian random effects

Number of observations            =   28091
Number of cases                   =    4697

X-var df          =    19
Sigma df          =     1
Scale df          =     1

Log likelihood =     -8774.6178     on   28070 residual degrees of freedom

Parameter               Estimate         Std. Err.
___________________________________________________
cons                     0.31033         0.12438
agebar                   0.33912E-01     0.89586E-02
age2bar                 -0.61971E-03     0.14671E-03
ttl_expbar               0.13992E-01     0.56496E-02
ttl_exp2bar              0.70780E-03     0.32449E-03
tenurebar                0.67410E-01     0.59389E-02
tenure2bar              -0.27607E-02     0.40271E-03
not_smsabar             -0.18732         0.10542E-01
southbar                -0.10128         0.95431E-02
agetilde                 0.35999E-01     0.33967E-02
age2tilde               -0.72299E-03     0.53421E-04
ttl_exptilde             0.33467E-01     0.29744E-02
ttl_exp2tilde            0.21627E-03     0.12813E-03
tenuretilde              0.35754E-01     0.18543E-02
tenure2tilde            -0.19701E-02     0.12537E-03
not_smsatilde           -0.89011E-01     0.95607E-02
southtilde              -0.60631E-01     0.10965E-01
grade                    0.61112E-01     0.19098E-02
black                   -0.60684E-01     0.98738E-02
sigma                    0.29158         0.13489E-02
scale                    0.24458         0.34461E-02
```

Discussion 3: Random effects models

The inference from the classical random effects model differs from that of the two extended random effects models. The inference from the two extended random effects models is the same. There is a significant difference between the likelihoods of the classical and extended random effects models, namely

$$-2(-8853.4259 - (-8774.6178)) = 157.62,$$

for $28078 - 28070 = 8$ degrees of freedom. Also several of the coefficients on the $\overline{x}_{pj}$ covariates are significant. This significance could be interpreted in two alternative ways: (1) the omitted effects are significantly correlated with the included time varying explanatory variables or (2) the explanatory variable time means have different impacts to their time-demeaned values.

12.3.6 Comparing 2-Level Fixed and Random Effects Models

As the FE results and the extended RE models make similar inferences about the effect of the time-varying covariates, it might seem that we can use either of them for inference about time-varying covariates. However, in this empirical comparison there were no internal covariates or state dependence effects, such as duration or lagged response. When these sorts of endogenous covariate are present, the correlation between the included and omitted effects will vary with time. This variation will depend on the current duration in a survival model (or the previous response in a first order model) and thus be difficult to capture in a fixed effects model.

In the absence of endogenous covariates, we can establish if there is some systematic non stationarity in the correlation between the included and omitted effects by dividing the observation window into blocks of responses and then producing time means and time-demeaned variable effects for each block.

To explore whether the coefficients for the time-constant covariates are really time-constant we can use dummy variables for different intervals of time and include the interactions of these dummy variables with the time-constant explanatory variables. However, it may not always be possible to account for the correlation between included covariates and the incidental parameters with simple linear functions of the means of the time-varying covariates, or by using different parameters for different intervals of time.

12.4 Fixed Effects Treatment of The 3-Level Linear Model

As we saw in Chapter 7, it is not unusual to have 3-level data, for instance, workers (level 2) in firms (level 3) employed over time (level 1). In the models discussed in Chapter 7, the lower level data were nested in their higher level units, and this simplified the analysis. However, with longitudinal data this 3-level nesting often gets broken, e.g. when workers change job and go to work for a different firm. When this happens, there is no transformation like time demeaning that will "sweep out" both the worker and firm fixed effects, see Abowd, Kramarz and Margolis (1999).

By focussing on different re-arrangements of the data (worker, firm and spell), different aspects of the model can be identified, e.g. the time-demeaned worker data identifies the differences in the firm effects for the workers who move, see Abowd, Creecy and Kramarz (2002). These different aspects of the model can then be recombined using minimum distance estimators, see Andrews et al (2006, 2008). Estimating the 3-level, linear model's fixed effects is particularly important for researchers who are interested in assessing their correlation with other effects in the model, e.g. Abowd et al. (1999, 2002) wanted to establish the relationship between "high wage workers and high wage firms".

12.5 Exercises

There are two exercises to accompany this section, namely: FE1 and FE2.

12.6 References

Abowd, J., Kramarz, F., and Margolis, D., (1999), High wage workers and high wage firms, Econometrica, 67, 251-333.

Abowd, J., Creecy, R. & Kramarz, F. (2002), Computing person and firm effects using linked longitudinal employer-employee data, Technical Paper 2002-06, U.S. Census Bureau, April.

Andersen, E.B., (1973), Conditional Inference and Models for Measuring, Copenhagen, Mentallhygiejnisk Forlag.

Andrews, M., Schank, T., and Upward, R., (2006), Practical fixed effects estimation methods for the three-way error components model, Stata Journal, 2006, 6, 461-481.

Andrews, M., Gill, L., Schank, T., and Upward, R., (2008), High wage workers and low wage firms: negative assortative matching or limited mobility bias?, forthcoming, Journal of the Royal Statistical Society Series A.

Chamberlain, G., (1980), Analysis of Covariance with Qualitative Data, Review of Economic Studies, 47, 225-238.

Cox, D.R., Reid, N., (1987), Parameter Othogonality and Approximate Conditional Inference, (with discussion), Journal of the Royal Statistical Society, B., 49, 1-39.

Crouchley, R., Pickles, A., (1989), An Empirical Comparison of Conditional and Marginal Likelihood Methods in a Longitudinal Study, Sociological Methodology, 19, 161-183.

Hsiao, C., (1986), Analysis of Panel Data, Cambridge University Press, Cambridge.

Hausman, J.A., (1978), Specification Tests in Econometrics, Econometrica, 46, 1251-1271.

Hausman, J. & Taylor, W. (1981), Panel data and unobservable individual effects, Econometrica, 49, 1377–98.

Heckman, J., (1981), The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time Stochastic Process, 179-197 in Structural Analysis of Discrete Data with Economic Applications, edited by C. Manski, McFadden, D., Cambridge, MA, MIT Press.

Mundlak, Y., (1978), On the Pooling of Time Series and Cross Sectional Data, Econometrica, 46, 69-85.

Neyman, J., Scott, E., (1948), Consistent Estimates Based on partially Consistent Observations, Econometrica, 46, 69-85.

Wooldridge, J. M. (2002), Econometric Analysis of Cross Section and Panel Data, MIT Press, Cambridge Mass.

Appendix A

Installation, Command Summary, Quadrature, Estimation and Endogenous Effects

A.1 Installation

For a Windows installation you will need to download the Sabre for Stata zip file from the Sabre site, `http://sabre.lancs.ac.uk/`, and unzip the files (1) `sabre_in_stata.plugin`, (2) `Sabre.ado`, to the sub directory `personal` of the directory `ado` on your C drive, so that Stata can find them. The files in the examples directory are the `do` and `dta` files for the examples and these need to go into your `Stata Data` directory. We have also included `do` and `dta` files for the exercises should you want them.

The file `sabre_in_stata.plugin` contains the C and FORTRAN code of the plugin. The file `Sabre.ado` contains the translation between the code needed by the plugin and the syntax we have used to delimit the sabre instructions in this book.

We have written `do` files for all the examples in this book, these are invoked interactively from within Stata using the Stata menu system (`File`>`do`), navigating to the directory where these `do` files and `dta` files are located and highlighting the `do` file we want to run.

To run the `sabre_in_stata.plugin` in batch mode with Stata on a Unix system you will need to use the format `stata --b do filename`, where 'filename' is the name of the `do` file

and not the format

`stata < filename.do.`

A.2 Sabre Commands

A.2.1 The Anatomy of a Sabre do file

There are various key elements to a Sabre do file. We will use the first few lines of the Poisson Model Example C5 (`prescribe.do`) to illustrate this; `prescribe.do` contains:

```
log using prescribe_s.log, replace
set more off
use racd
#delimit ;
sabre, data sex age agesq income levyplus freepoor freerepa illness actdays
hscore chcond1 chcond2 dvisits nondocco hospadmi hospdays medicine
prescrib nonpresc constant id;
sabre sex age agesq income levyplus freepoor freerepa illness actdays hscore
chcond1 chcond2 dvisits nondocco hospadmi hospdays medicine prescrib
nonpresc constant id, read;
#delimit cr
sabre, case id
sabre, yvar prescrib
sabre, family p
sabre, constant cons
#delimit ;
sabre, lfit sex age agesq income levyplus freepoor freerepa illness actdays
hscore chcond1 chcond2 cons;
#delimit cr
sabre, dis m
sabre, dis e
#delimit ;
sabre, fit sex age agesq income levyplus freepoor freerepa illness actdays
hscore chcond1 chcond2 cons;
#delimit cr
sabre, dis m
sabre, dis e
log close
clear
exit
```

We will now break this down into various parts and explain what each part does.

The following commands are Stata commands which open a log file called `prescribe_s.log`, and load the data set `racd.dat` into Stata. Remember that Sabre is a plugin and not actually part of Stata.

```
log using prescribe_s.log, replace
set more off
use racd
```

The following commands are used to transfer the variables we want to use in the model fitting from Stata into Sabre

```
#delimit ;
sabre, data sex age agesq income levyplus freepoor freerepa illness actdays
hscore chcond1 chcond2 dvisits nondocco hospadmi hospdays medicine
prescrib nonpresc constant id;
sabre sex age agesq income levyplus freepoor freerepa illness actdays hscore
chcond1 chcond2 dvisits nondocco hospadmi hospdays medicine prescrib
nonpresc constant id, read;
#delimit cr
```

The command below tells Sabre what the level-2 (`case`, grouping variable) is called, in this example it is `id`.

```
sabre, case id
```

The following commands tell Sabre what the response variable is called, in this case its `prescrib`. The commands then tell Sabre which member of the exponential family we want to use, in this case the `p` is for the Poisson model,. The last command tells Sabre that that we want to include a constant (`cons`) in the model.

```
sabre, yvar prescrib
sabre, family p
sabre, constant cons
```

The following commands (`lfit`) tell Sabre to estimate a non random effects model (of the covariates `sex-chcond2` and a constant (`cons`))..and then display (`dis`) the model (`m`) and the estimates (`e`) of this fit.

```
#delimit ;
sabre, lfit sex age agesq income levyplus freepoor freerepa illness actdays
hscore chcond1 chcond2 cons;
#delimit cr
sabre, dis m
sabre, dis e
```

The following commands (`fit`) tell Sabre to estimate a random effects model (of the covariates `sex-chcond2` and a constant `(cons)`).using the default quadrature (normal) with the default number of mass points (12) and then display (`dis`) the model (`m`) and the estimates (`e`) of this fit.

```
#delimit ;
sabre, fit sex age agesq income levyplus freepoor freerepa illness actdays
hscore chcond1 chcond2 cons;
#delimit cr
sabre, dis m
sabre, dis e
```

The following are Stata commands that `close` the log file, `clear` Stata's memory and `exit` Stata.

```
log close
clear
exit
```

A.2.2 Command Summary

The list of variable names is declared, and the dataset read into sabre using:

```
use filename(.dta)

sabre, data variable_list

sabre variable_list, read
```

Long lines in a .do file can be written using:

```
#delimit ;
```

sabre_in_stata command written on more than one line;

```
#delimit cr
```

For all model types, a homogeneous model is fitted using:

```
sabre, LFIT variable_list
```

The fitted model and the parameter estimates are displayed using:

```
sabre, DISPLAY m

sabre, DISPLAY e
```

A.2.3 Random effects models

For all model types, a random effects model is fitted using:

```
sabre, FIT variable_list
```

Univariate models

```
sabre, CASE variable_name

sabre, YVARIATE variable_name

sabre, FAMILY b (default) or p or g

sabre, LINK l (default) or p or c if FAMILY b

sabre, ENDPOINT n (default) or l or r or b if FAMILY b

sabre, ENDPOINT n (default) or l if FAMILY p

(sabre, CONSTANT variable_name)

sabre, QUADRATURE g (default) or a

sabre, MASS positive integer (default is 12)
```

Univariate ordered response models

```
sabre, CASE variable_name

sabre, YVARIATE variable_name

sabre, ORDERED y

sabre, LINK l (default) or p

sabre, QUADRATURE g (default) or a

sabre, MASS positive integer (default is 12)
```

Bivariate models

```
sabre, CASE variable_name

sabre, YVARIATE variable_name

sabre, MODEL b

sabre, RVARIATE variable_name

sabre, FAMILY FIRST = b (default) or p or g SECOND = b (default) or p or g
```

```
sabre, LINK FIRST = l (default) or p or c if FAMILY FIRST = b

 SECOND = l (default) or p or c if FAMILY SECOND = b

(sabre, CONSTANT FIRST = variable_name SECOND = variable_name)

sabre, NVAR positive integer

sabre, MASS FIRST = positive integer (default is 12)

 SECOND = positive integer (default is 12)
```

Trivariate models

```
sabre, CASE variable_name

sabre, YVARIATE variable_name

sabre, MODEL t

sabre, RVARIATE variable_name

sabre, FAMILY FIRST = b (default) or p or g SECOND = b (default) or p or g

 THIRD = b (default) or p or g

sabre, LINK FIRST = l (default) or p or c if FAMILY FIRST = b

 SECOND = l (default) or p or c if FAMILY SECOND = b

 THIRD = l (default) or p or c if FAMILY THIRD = b

(sabre, CONSTANT FIRST = variable_name SECOND = variable_name

 THIRD = variable_name)

sabre, NVAR FIRST = positive integer SECOND = positive integer

sabre, QUADRATURE g (default) or a

sabre, MASS FIRST = positive integer (default is 12)

 SECOND = positive integer (default is 12)

 THIRD = positive integer (default is 12)
```

Univariate first order models

```
sabre, CASE variable_name

sabre, YVARIATE variable_name
```

```
sabre, DEPEND y

sabre, RVARIATE variable_name

sabre, FAMILY b (default) or p or g

sabre, LINK l (default) or p or c if FAMILY b

(sabre, CONSTANT variable_name)

sabre, NVAR positive integer

sabre, QUADRATURE g (default) or a

sabre, MASS positive integer (default is 12)
```

Bivariate first order models

```
sabre, CASE variable_name

sabre, YVARIATE variable_name

sabre, MODEL b

sabre, RVARIATE variable_name

sabre, EQSCALE y

sabre, DER1 y

sabre, FAMILY FIRST = b (default) or p or g, SECOND = b (default) or p or g

sabre, LINK FIRST = l (default) or p or c if FAMILY FIRST = b

 SECOND = l (default) or p or c if FAMILY SECOND = b

(sabre, CONSTANT FIRST = variable_name SECOND = variable_name)

sabre, NVAR positive integer

sabre, QUADRATURE g (default) or a

sabre, MASS FIRST = positive integer (default is 12)

 SECOND = positive integer (default is 12)
```

Multilevel models

```
sabre, CASE FIRST = variable_name, SECOND = variable_name

sabre, YVARIATE variable_name
```

```
sabre, FAMILY b (default) or p or g

sabre, LINK l (default) or p or c if FAMILY b

(sabre, CONSTANT variable_name)

sabre, MASS FIRST = positive integer (default is 12)

 SECOND = positive integer (default is 12)
```

A.2.4 Fixed effects models

For all model types, a fixed effects model is fitted using:

```
sabre, FEFIT variable_list
```

Univariate models (linear model only)

```
sabre, CASE variable_name

sabre, YVARIATE variable_name

sabre, FAMILY g

(sabre, CONSTANT variable_name)
```

A.2.5 Other Sabre Commands

```
sabre, ALPHA real, positive; 0.01

sabre, APPROX integer, non-negative <= MAXIMU; 5

sabre, ARITHM character, 'f(ast)' or 'a(ccurate)'; 'f'

sabre, CASE variable name(s) [FIRST=, SECOND=]; FIRST is first name listed in

 DATA command

sabre, COMMEN text

sabre, CONSTA variable name(s) [FIRST=, SECOND=, THIRD=]

sabre, CONVER real, positive; 5e-5

sabre, CORREL character, 'y(es)' or 'n(o)'; 'y'

sabre, CUTPOI reals, monotonically increasing
```

```
sabre, DATA list of variable names

sabre, DEFAUL no argument

sabre, DELETE list of variable names

sabre, DEPEND character, 'y(es)' or 'n(o)'; 'n'

sabre, DER1 character, 'y(es)' or 'n(o)'; 'n'

sabre, DISPLA character, 'e(stimates)' or 'l(imits)' or 'm(odel)' or
 's(ettings)' or 'v(ariables)'

sabre, ENDPOI character, 'b(oth)' or 'l(eft)' or 'r(ight)' or 'n(one)'; 'n'

sabre, EQSCAL character, 'y(es)' or 'n(o)'; 'n'

sabre, FACTOR variable name, <space>, categorical variable name

sabre, FAMILY character(s), 'b(inomial)' or 'g(aussian)' or 'p(oisson)'
 [FIRST=, SECOND=, THIRD=]; 'b'

sabre, FEFIT list of variable names

sabre, FIT list of variable names

sabre, FIXRHO no argument, or real, between -1 and +1

sabre, FIXSC1 no argument, or real, positive

sabre, FIXSCA no arguments, or reals, positive [FIRST=, SECOND=]

sabre, HELP no argument

sabre, HISTOG variable name (integer, between 2 and 11)

sabre, INITIA list of reals

sabre, INPUT filename

sabre, LFIT list of variable names

sabre, LINK character, 'l(ogit)' or 'p(robit)' or 'c(omplementary log-log)'
 [FIRST=, SECOND=, THIRD=]; 'l'

sabre, LOOK list of variable names (integer, positive, <space>, integer,
 positive)

sabre, MASS integer, 2,4,...,16,20,...,48,56,...,112,128,...,256; 12

sabre, MAXIMU integer, positive >= APPROX; 100
```

```
sabre, MODEL character, 'u(univariate)' or 'b(ivariate)' or 't(rivariate)';

 'u'

sabre, NVAR integer, positive [FIRST=, SECOND=]

sabre, OFFSET variable name

sabre, ORDERE character, 'n(o)' or 'y(es)'; 'n'

sabre, OUTPUT filename; "sabre.log"

sabre, QUADRA character, 'g(aussian)' or 'a(daptive)'; 'g'

sabre list of variable names (as in 'sabre, DATA' command), READ

sabre, RESID filename; "sabre.res"

sabre, RESTAR no argument

sabre, RHO real, between -1 and +1 [FIRST=, SECOND=, THIRD=]

sabre, RVARIA variable name

sabre, SCALE real, positive [FIRST=, SECOND=, THIRD=]

sabre, SIGMA real, positive [FIRST=, SECOND=, THIRD=]

sabre, STOP no argument

sabre, TIME filename; "sabre.time"

sabre, TOLERA real, positive; 1e-6

sabre, TRACE filename; "sabre.trace"

sabre, TRANSF (various forms of syntax)

sabre, YVARIA variable name
```

A.3 Quadrature

We illustrate normal Gaussian quadrature and adaptive Gaussian quadrature for the 2-level Generalised Linear Model (GLM). The ideas can be extended to higher levels and to multivariate responses.

The 2 level GLM likelihood takes the form

$$L\left(\gamma, \phi, \sigma_{u_0}^2 | \mathbf{y}, \mathbf{x}, \mathbf{z}\right) = \prod_j \int_{-\infty}^{+\infty} \prod_i g\left(y_{ij} \mid \theta_{ij}, \phi\right) f\left(u_{0j}\right) du_{0j},$$

where

$$g\left(y_{ij} \mid \theta_{ij}, \phi\right) = \exp\left\{\left[y_{ij}\theta_{ij} - b\left(\theta_{ij}\right)\right]/\phi + c\left(y_{ij}, \phi\right)\right\},$$

$$\theta_{ij} = \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + u_{0j},$$

and

$$f\left(u_{0j}\right) = \frac{1}{\sqrt{2\pi}\sigma_{u_0}} \exp\left(-\frac{u_{0j}^2}{2\sigma_{u_0}^2}\right).$$

Sabre can evaluate the integrals in $L\left(\gamma, \phi, \sigma_{u_0}^2 | \mathbf{y}, \mathbf{x}, \mathbf{z}\right)$ for the multilevel GLM model using either normal Gaussian or adaptive Gaussian quadrature.

A.3.1 Normal Gaussian Quadrature

Normal Gaussian quadrature or just Gaussian Quadrature uses a finite number (C) of quadrature points consisting of weights (probabilities= p_c) and locations u_0^c. The values of p_c and u_0^c are available from standard normal tables, e.g. Stroud and Secrest (1966). The approximation takes the form

$$L\left(\gamma, \phi, \sigma_{u_0}^2 | \mathbf{y}, \mathbf{x}, \mathbf{z}\right) \simeq \prod_j \sum_{c=1}^{c=C} p_c \prod_i g\left(y_{ij} \mid \theta_{ij}^c, \phi\right),$$

where

$$\theta_{ij}^c = \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + \sigma_{u_0} u_0^c,$$

$$\sum_{c=1}^{c=C} p_c = 1.$$

The approximation works so long as $\prod_i g\left(y_{ij} \mid \theta_{ij}, \phi\right)$ can be represented by a polynomial in u_{0j} which is of degree less than or equal to $2C-1$. However, it is not a priori clear what value of C is required. Consequently, it is important to check whether enough quadrature points have been used by comparing solutions. Typically we start with a small C and increase it until covergence in the likelihood occurs. When C is large enough, the addition of more quadrature points wont improve the approximation.

Sabre can use: 2, (2),16; 16,(4),48; 48,(8),112; 112,(16),256 quadrature points for each random effect. The notation a,(b),c means from a to c in steps of length b. In Stata and gllamm the number of quadrature points must be between 4 and 195.

A.3.2 Performance of Gaussian Quadrature

In serial Sabre (as distinct from parallel Sabre) the larger the number of quadrature points used, the longer it takes to compute the likelihood. The time taken is roughly proportional to the product of the number of quadrature points for all the random effects in the multivariate multilevel GLM. For a bivariate 2-level random intercept model there are two random effects at level-2 for each response. If we use $C = 16$ quadrature points for each random effect, then the total time will be approximately $16^2 = 256$ times longer than a model without any random effects ($C = 1$).

Rabe-Hesketh, et al (2005) noted that Gaussian quadrature (or Normal Quadrature (NQ)) tends to work well with moderate cluster sizes as typically found in panel data. However with large cluster sizes, which are common in grouped cross-sectional data, the estimates from some algorithms can become biased. This problem was articulated by Borjas and Sueyoshi (1994) and Lee (2000) for probit models, by Albert and Follmann (2000) for Poisson models and by Lesaffre and Spiessens (2001) for logit models.

Lee (2000) attributes the poor performance of quadrature to numerical underflow and develops an algorithm to overcome this problem.

Rabe-Hesketh et al (2005) noted that for probit models the Lee (2000) algorithm works well in simulations with clusters as large as 100 when the intraclass correlation is 0.3 but produces biased estimates when the correlation is increased to 0.6. Rabe-Hesketh et al (2005) note that a likely reason for this is that for large clusters and high intraclass correlations, the integrands of the cluster contributions to the likelihood have very sharp peaks that may be located between adjacent quadrature points.

There can be problems with underflow and overflow in Sabre when estimating models. if this occurs, Sabre will give you a warning message and suggest you use a more accurate form of arithmetic. In some contexts the undeflow can be benign, for instance when we calculate

$$p_c \prod_i g\left(y_{ij} \mid \theta_{ij}^c, \phi\right),$$

for the tails of the distribution, the contribution to the total can be so close to zero, it will makes no real difference to the total (sum over c) and can be ignored.

By default, Sabre uses standard double precision (FORTRAN 95, real*8) variables and arithmetic (`sabre, ARITHM f(ast)`). This is adequate for most applications but occasionally, some of the intermediate calculations of the log likelihood $\log L\left(\gamma, \phi, \sigma_{u_0}^2 | \mathbf{y}, \mathbf{x}, \mathbf{z}\right)$, and its 1st and 2nd order derivatives can require the calculation of values which are beyond the range of double precision numbers. This range is approximately 10 to the power -308 to 10 to the power +308.

This range can be greatly extended by using the command `sabre, ARITHM` `a(ccurate)`. In this case all calculations are performed using specially written arithmetic code in which the exponent of the variable is stored separately in a 4 byte integer. This extends the range of intermediate calculations to approximately 10 to the power -2 billion to 10 to the power +2 billion. The precision with which numbers are stored is the same for both `'f(ast)'` and `'a(ccurate)'`, viz. about 15 decimal digits.

The greater range comes at the cost of increased run time, typically 15 times as long as in `fast` arithmetic. However, particularly when using parallel Sabre on a large number of processors, this may be a cost well worth paying as the problem may not otherwise be soluble. Neither Stata nor SAS have an equivalent to Sabre's `'a(ccurate)'` procedure.

By default Sabre uses normal Gaussian quadrature (`sabre, QUADRATURE g)`. Rabe-Hesketh et al (2005) proposed the use of adaptive quadrature as an alternative to normal quadrature (g), partly to avoid the problem of underflow/overflow that occurs in Normal Gaussian Quadrature. Adaptive Quadrature will be performed by Sabre if you use the command `sabre, QUADRATURE a`.

A.3.3 Adaptive Quadrature

Adaptive Quadrature works by adapting the quadrature locations of each integral inorder to place them where they are of most benefit to the quadrature approximation, i.e. under the peaks. The adaptive quadrature weights and locations depend on the parameters of the model. Between each step of the maximization algorithm the weights and locations are shifted and rescaled. We follow Skrondal and Rabe-Hesketh (2004), Rabe-Hesketh et al (2005) in illustrating AQ. If we adopt a Bayesain perpective, i.e assume that we know the model parameters $\left(\gamma, \phi, \sigma_{u_0}^2\right)$, then the the 2 level GLM likelihood

$$L\left(\gamma, \phi, \sigma_{u_0}^2 | \mathbf{y}, \mathbf{x}, \mathbf{z}\right) = \prod_j \int_{-\infty}^{+\infty} \prod_i g\left(y_{ij} \mid \theta_{ij}, \phi\right) f\left(u_{0j}\right) du_{0j},$$

has an integrand that is made up of the product of the joint probabilty of the data given u_{0j} and the prior density of u_{0j}, i.e.

$$\prod_i g\left(y_{ij} \mid \theta_{ij}, \phi\right) f\left(u_{0j}\right).$$

Under the Bayesian central limit theorem (Carlin and Louis 2000, p122-124), posterior densities are approximately normal. If μ_j, φ_j^2 are the mean and variance of this posterior density $f\left(u_{0j}; \mu_j, \varphi_j^2\right)$, then the ratio

$$\frac{\prod_i g\left(y_{ij} \mid \theta_{ij}, \phi\right) f\left(u_{0j}\right)}{f\left(u_{0j}; \mu_j, \varphi_j^2\right)},$$

should be approximated by a lower degree polynomial than the orginal Gaussain quadrature function. (If this is the case we will require fewer quadrature points than normal Gaussian quadrature.) We can rewrite the orginal GQ integral as

$$f_j\left(\gamma,\phi,\sigma_{u_0}^2\right)=\int_{-\infty}^{+\infty} f\left(u_{0j};\mu_j,\varphi_j^2\right)\left[\frac{\prod_i g\left(y_{ij}\mid\theta_{ij},\phi\right)f\left(u_{0j}\right)}{f\left(u_{0j};\mu_j,\varphi_j^2\right)}\right]du_{0j},$$

so that we the posterior density $f\left(u_{0j};\mu_j,\varphi_j^2\right)$ becomes the weight function. Let $f\left(\nu_j\right)$ denote a standard normal density, then by applying the change of variable

$$\nu_j=\frac{\left(u_{0j}-\mu_j\right)}{\varphi_j},$$

to the elements of $f_j\left(\gamma,\phi,\sigma_{u_0}^2\right)$, and applying the standard quadrature rule (with weights p_c and locations ν_0^c), θ_{ij}^c becomes

$$\theta_{ij}^{AQc}=\gamma_{00}+\sum_{p=1}^{P}\gamma_{p0}x_{pij}+\sum_{q=1}^{Q}\gamma_{0q}z_{qj}+\sigma_{u_0}\left(\varphi_j\nu_0^c+\mu_j\right),$$

and

$$\begin{aligned}f_j\left(\gamma,\phi,\sigma_{u_0}^2\right)&\simeq\sum_c p_c\left[\frac{\prod_i g\left(y_{ij}\mid\theta_{ij}^{AQc},\phi\right)f\left(\varphi_j\nu_0^c+\mu_j\right)}{\frac{1}{\varphi_j\sqrt{2\pi}}\exp-\left(\nu_0^c\right)^2}\right]\\&=\sum_c\pi_{jc}\left[\prod_i g\left(y_{ij}\mid\theta_{ij}^{AQc},\phi\right)\right],\end{aligned}$$

where

$$\pi_{jc}=p_c\left[\frac{f\left(\varphi_j\nu_0^c+\mu_j\right)}{\frac{1}{\varphi_j\sqrt{2\pi}}\exp-\left(\nu_0^c\right)^2}\right].$$

Unfortunately, at each interation of the optimization procedure, the posterior mean and variance $\left(\mu_j,\varphi_j^2\right)$ of each group are not apriori known. They can however be obtained from an iterative procedure, see Naylor and Smith (1988). Let us use the superscript k to denote values at the k th iteration. At the start we have $k=1$, and set $\mu_j^0=0$, and $\varphi_j^0=1$, to give $\varphi_j^0\nu_0^c+\mu_j^0$ and π_{jc}^0. The posterior means and variances are then updated at each subsequent iteration using

$$\mu_j^k=\frac{\sum_c\left(\varphi_j^{k-1}\nu_0^c+\mu_j^{k-1}\right)\pi_{jc}^{k-1}\left[\prod_i g\left(y_{ij}\mid\theta_{ij}^{AQc^{k-1}},\phi^{k-1}\right)\right]}{f_j^k\left(\gamma^k,\phi^k,\sigma_{u_0}^{2k}\right)},$$

$$\left(\varphi_j^k\right)^2=\frac{\sum_c\left(\varphi_j^{k-1}\nu_0^c+\mu_j^{k-1}\right)^2\pi_{jc}^{k-1}\left[\prod_i g\left(y_{ij}\mid\theta_{ij}^{AQc^{k-1}},\phi^{k-1}\right)\right]}{f_j^k\left(\gamma^k,\phi^k,\sigma_{u_0}^{2k}\right)}-\left(\mu_j^k\right)^2,$$

where

$$f_j^k\left(\gamma^k,\phi^k,\sigma_{u_0}^{2^k}\right)\simeq\sum_c\pi_{jc}^{k-1}\left[\prod_i g\left(y_{ij}\mid\theta_{ij}^{AQc^{k-1}},\phi^{k-1}\right)\right].$$

At each k, we use μ_j^{k-1}, and φ_j^{k-1} in $\varphi_j^{k-1}\nu_0^c+\mu_j^{k-1}$ and π_{jc}^k, to start the convergence process that will give us μ_j^k, and φ_j^k. As the optimization procedure gets closer to the solution, there is less change in $\gamma^k,\phi^k,\sigma_{u_0}^{2^k}$, and consequently in μ_j^k, and φ_j^k, so that convergence in this local adaptation occurs in 2-3 cycles.

It is our experience that underflow can still occur in Sabre with Adaptive Quadrature (`sabre, QUADRATURE a`) but this can be resolved by using the command,.`sabre, ARITHM a(ccurate).` Algorithms for Adaptive Quadrature for multilevel and multivariate random effects can also be developed along similar lines, see Skrondal and Rabe-Hesketh (2004), Rabe-Hesketh et al (2005). Adaptive Quadrature has been deployed in Sabre for univariate, bivariate and trivariate 2 level GLMs. However, it has not yet been deployed for 3 level models.

A.4 Estimation

Two forms of estimation are considered: (1) Random Effect Models, (2) Fixed Effect Models.

A.4.1 Maximizing the Log Likelihood of Random Effect Models

Sabre uses the Newton-Raphson algorithm to maximize the log-likelihood. The Newton-Raphson algorithm is an iterative procedure. If we denote the parameters $\left(\pi=\gamma,\phi,\sigma_{u_0}^2\right)$ which maximize $\log L\left(\pi|\mathbf{y},\mathbf{x},\mathbf{z}\right)$, then a necessary condition for this to occur is

$$\frac{\partial\log L\left(\pi|\mathbf{y},\mathbf{x},\mathbf{z}\right)}{\partial\pi}=0.$$

If we let the values of the parameters at the nth iteration be denoted by π^n. Then a 1st order Taylor expansion about π^n gives

$$\begin{aligned}\frac{\partial\log L\left(\pi|\mathbf{y},\mathbf{x},\mathbf{z}\right)}{\partial\pi}&\simeq\left[\frac{\partial\log L\left(\pi|\mathbf{y},\mathbf{x},\mathbf{z}\right)}{\partial\pi}\right]_{\pi=\pi^n}\\&+\left[\frac{\partial^2\log L\left(\pi|\mathbf{y},\mathbf{x},\mathbf{z}\right)}{\partial\pi\partial\pi'}\right]_{\pi=\pi^n}\left(\pi-\pi^n\right)\\&=g\left(\pi^n\right)+H\left(\pi^n\right)\left(\pi-\pi^n\right),\end{aligned}$$

where $g\left(\pi^n\right)$ is the gradient vector at π^n and $H\left(\pi^n\right)$ is the Hessian. The process is made iterative by writing

$$g\left(\pi^n\right)+H\left(\pi^n\right)\left(\pi^{n+1}-\pi^n\right)=0,$$

so that

$$\pi^{n+1} = \pi^n - \left[H\left(\pi^n\right)\right]^{-1} g\left(\pi^n\right).$$

When π has say k elements, $(k > 1)$, the computational effort required to calculate $\log L\left(\pi^n|\mathbf{y},\mathbf{x},\mathbf{z}\right)$ once is much less than it is to calculate $g\left(\pi^n\right)$ for the k elements of π and similarly for the $(k-1)k/2$ distinct elements of $H\left(\pi^n\right)$. So we actually use

$$\begin{aligned}\pi^{n+1} &= \pi^n + s\left[-H\left(\pi^n\right)\right]^{-1} g\left(\pi^n\right) \\ &= \pi^n + s\mathbf{d},\end{aligned}$$

where s is a scalar (often called the step length). At each step (n) we try $s = 1$, if

$$\log L\left(\pi^{n+1}|\mathbf{y},\mathbf{x},\mathbf{z}\right) \succ \log L\left(\pi^n|\mathbf{y},\mathbf{x},\mathbf{z}\right),$$

then continue. While if

$$\log L\left(\pi^{n+1}|\mathbf{y},\mathbf{x},\mathbf{z}\right) \preceq \log L\left(\pi^n|\mathbf{y},\mathbf{x},\mathbf{z}\right),$$

try $s = 0.5$, or, $s = 0.25$, untill

$$\log L\left(\pi^{n+1}|\mathbf{y},\mathbf{x},\mathbf{z}\right) \succ \log L\left(\pi^n|\mathbf{y},\mathbf{x},\mathbf{z}\right),$$

then continue.

Sabre also has an option that allows you to use minus the outer product of the gradient vectors, which we write as

$$H\left(\pi^n\right) = -\sum\nolimits_j g_j\left(\pi^n\right) g_j\left(\pi^n\right)'.$$

In the 2 level GLM $g_j\left(\pi^n\right)$ takes the form

$$g_j\left(\pi^n\right) = \frac{\partial\left[\log\sum_{c=1}^{c=C} p_c \prod_i g\left(y_{ij} \mid \theta_{ij}^c, \phi\right)\right]_{\pi^n}}{\partial\pi}.$$

The outer product of the gradient vectors ensures that $H\left(\pi^n\right)$ is negative definite, this form of $H\left(\pi^n\right)$ can be useful when there are many local maxima and minima of $\log L\left(\pi|\mathbf{y},\mathbf{x},\mathbf{z}\right)$. This version of $H\left(\pi^n\right)$ gives the Fisher-Scoring algorithm, see Berndt et al (1974), however, it can be very slow to converge when compared to Newton-Raphson algorithm for estimating mutivariate multilevel GLMs (evaluated using Gaussian quadrature).

It is important to acknowledge that many Gaussian quadrature loglikelihoods have multiple local maxima, this makes it necessary to use different starting values, compare the solutions and establish the best. It is only the global maxima in $\log L\left(\pi|\mathbf{y},\mathbf{x},\mathbf{z}\right)$ that provides the maximum likelihood estimates.

Sabre uses analytic rather than numerical approximations to $H\left(\pi^n\right)$ and $g\left(\pi^n\right)$. This makes Sabre much faster than gllamm (Stata) which uses `ml` (Newton-Raphson) with method `d0` (no analytic derivaties required).

A.4.2 Fixed Effect Linear Models

Using the notation of Chapter 3 and 12 for the linear model, the explanatory variables at the individual level are denoted by $x_1, \cdots, x_P$, and those at the group level by $z_1, \cdots, z_Q$, so that

$$y_{ij} = \gamma_{00} + \sum_{p=1}^{P} \gamma_{p0} x_{pij} + \sum_{q=1}^{Q} \gamma_{0q} z_{qj} + u_{0j} + \varepsilon_{ij}.$$

The regression parameters γ_{p0} $(p = 1, \cdots, P)$ and γ_{0q} $(q = 1, \cdots, Q)$ are for level-one and level-two explanatory variables, respectively. Groups with only one individual have to be removed from the data before the data is processed, as the dummy variables for groups of size 1 are not identified. This model is estimated without a constant and time constant covariates, i.e. we set $\gamma_{00} = \gamma_{0q} = 0$, and treat all of the incidental parameters u_{0j} as dummy variables. This is the Least Squares Dummy Variable (LSDV) estimator, the estimates of u_{0j} are biased but consistent. A number of fixed effects estimators have been proposed, we use the term LSDV for the explicit use of dummy variables. For some other papers on the estimation of this model see e.g. Abowd et al (2002), Andrews et al (2005).

There can be too many groups in a data set to perform the conventional matrix manipulations needed to estimate this model in the limited memory of most desktop PCs. Sabre does not use any approximations or differencing (demeaning), as it directly solves the least squares normal equations for the model. Further, the group sizes do not need to be balanced. The algorithm still works if the model includes level 3 (dummy variables) so long as they change for some level 2 subjects. To solve the normal equations Sabre uses some of the large sparse matrix algorithms of the Harwell Subroutine Library (HSL), see `http://www.cse.scitech.ac.uk/nag/hsl/`. The Sabre estimator `(sabre, FEFIT variable_list)` also goes parallel on multiprocessor systems.

A.4.3 The Relative Performance of Different Software Packages for Estimating Multilevel Random Effect Models

In this section we compare Sabre, Stata, gllamm (Stata), for estimating multivariate multilevel random effect models on Lancaster's HPC. . In all comparisons we use the default or recommended starting values of the different procedures. The HPC execution nodes are 124 Sun Fire X4100 servers with two dual-core 2.4GHz Opteron CPUs, for a total of 4 CPUs per node. The standard memory per node is 8G, with a few nodes offering 16G. Most nodes also offer dedicated inter-processor communication in the form of SCore over gigabit Ethernet, to support message passing (parallel) applications.

Example	Data	Obs	Vars	Kb	Stata	gllamm	Sabre(1)	Sabre(8)
L7	filled	390432	94	367556	59hr 52'	3 months+	34' 38"	7' 01"
	lapsed	390432	94	367556	67hr 31'	3 months+	29' 41"	6' 21"
L8	filled-lapsed	780864	261	2134413	n/a	3 years+	54hr 29'	11hr 58'
L9	union-wage	37990	25	9683	n/a	unexpected failure	18' 21"	2' 26"

Key

Sabre (8) Is Sabre running on 8 processors

n/a: Stata 9 can not estimate bivariate random effects models using quadrature

unexpected failure: the gllamm manual not does rule this bivariate model out, but gllamm crashed just after starting

time+: indicates a lower CPU limit

Obs is the number of Observations in the data set

Vars is the number of explanatory variables in the model

Kb is the size of the raw data set in kilobytes.

The Example L7 (filled) is for a random effects complementary log log link model of duration with a non parametric baseline hazard.(by week) until filled of a job vacancy using micro data. Example L7 (lapsed) is the same, except this time the duration (by week) is until the vacancy is lapsed, i.e. withdrawn from the market by the firm. The Example L8 is the joint analysis of the filled and lapsed data using a bivariate random effects competing risk model with non parametric baseline hazards of each failure type. For further details on these classes of model see Chapter 9, Event History Models. The data are from Lancashire Careers Service over the period 1985-1992, for a copy of the data see `http://sabre.lancs.ac.uk/comparison3.html.` We consider Examples L7 and L8 to be medium to large sized data sets.

The Example L9 is the NLS wage panel example of adult males from the Stata manual. The model is a bivariate random effects model of wages (linear model) and trade union membership (binary response) sequences, in which trade union membership also has a direct effect on wages. For further details on this class of model see Chapter 8, Multivariate 2 Level Generalised Linear Models. We consider Example L9 to be a small data set.

It is not just the actual times that are important, as these will get shorter on faster processors, though perhaps not fast enough to enable this type of analysis on large data sets for many years to come. Researchers in the exploratory phase of their work typically need quick but reliable estimates of the many different versions of the model they are trying to fit.

For the relative times, this Table shows that serial Sabre(1) is over 100 times faster than Stata and about 3000 times faster than gllamm on the univariate data sets (L7). When we go parallel with Sabre we get another 5 fold increase in speed on 8 processors. For the bivariate Example L8, we estimate that Sabre (1) is at least 500 times faster than gllamm and that Sabre has a further 5 fold increase in speed on 8 processors. These are crude estimates as we had to estimate the termination time for gllamm from the few iterations that it was able to produce after 2 weeks running. Neither Stata (9) nor Stata (10) are able to estimate bivariate random effect models using quadrature

What is scientifically important, is that our substantive results change as the modelling becomes more comprehensive. For instance, the bivariate analysis (L8) not only provided a estimate of the correlation in the unobserved Random effects of the different vacancy types but the inference on a range of explanatory variables changed (compared to that of L7). In the NLS data the coefficient of trade union membership in the wage equation is much smaller in the bivariate model, when compared to that obtained when we assume that the sequence of wages is independent from the sequence of trade union memberships. In other words, the impact of trade union membership on wages is not as large as the analsysis that assumes indpendence would suggest.

For further comparisons on other small and medium sized data sets and with other software systems see `http://sabre.lancs.ac.uk/comparison3.html`. The bigger the data set, or the more complex the model, the better the relative performance of Sabre. In all comparisons the numerical properties of Sabre's estimates compare favourably with those of the alternatives and it has the best overall computational speed. The speed produced by Sabre make it possible to explore many more comprehensive model specifications in a reasonable time period.

A.5 Endogenous and Exogenous Variables

In the social sciences, interest often focuses on the dynamics of social or economic processes. Social science theory suggests that individual behaviour, choices or outcomes of a process are directly influenced by (or are a function of) previous behaviour, choices or outcomes. For instance, someone employed this week is more likely to be in employment next week than someone who is currently unemployed; someone who voted for a certain political party in the last elections is more likely to vote for that party in the next elections than someone who did not.

When analysing observational data in the social sciences, it is necessary to distinguish between two different types of explanatory variable; those which are exogenous (or external) to the process under study (for example age, sex, social class and education in studies of voting behaviour), and those which are endogenous . Endogenous variables have characteristics which in some way relate to previous decisions, choices or outcomes of a process. For example, in a study of voting behaviour previous vote, being a previous decision, is an endogenous variable; in the study of migration, duration of stay since the last residential move is endogenous as it relates to previous migration behaviour.

Endogenous variables may be seen as proxy variables for the many unmeasured and unmeasurable factors which affect individual choice or behaviour and which are therefore necessarily omitted from analyses. Thus voting choice may be seen as a proxy for individual social, economic and psychological characteristics, while duration of stay in a locality is a proxy for all the unknown social and economic factors which affect an individual's propensity to move.

Endogenous variables create problems in statistical analyses, because being related to the outcomes of the process of interest they will, by definition, be a function of the unobserved variables which govern the process. They will therefore be correlated with the random variation (or error structure) of the outcome. This leads to an infringement of the basic regression model assumption that the explanatory variables included in the model are independent of the error term. The consequence of this violation is risk of substantial and systematic bias.

In the presence of endogenous variables the basic statistical models are not robust against the infringement of assumptions. Expressed technically, parameter estimation is not consistent, ie. there is no guarantee that the parameter estimates will approach their true values as the sample size increases. Consistency is usually regarded as the minimum requirement of an acceptable estimation procedure.

To avoid spurious relationships and misleading results, with endogenous variables it is essential to use longitudinal data and models in which there is control for omitted variables. Longitudinal data, and in particular repeated measures on individuals are important because they provide scope for controlling for individual specific variables omitted from the analysis.

The conventional approach to representing the effect of omitted variables is to add an individual specific random term to the linear predictor, and to include an explicit distribution for this random term in the model.

There is no single agreed terminology for models which include this random tem. In econometrics the models are called random effect models; in epidemiology, frailty models; and statisticians also refer to them as multilevel models, mixture models or heterogeneous models. Models without random effects are sometimes called homogeneous models. An alternative terminology describes models without random effects as marginal models and models with random effects as conditional models. Marginal models correspond closely to the "population averaged" formulations used in the General Estimating Equation literature.

It is important to note that when interest focuses on the causal relationship in social processes inference can only be drawn by using longitudinal data and models in which there is control for unobserved (or residual) heterogeneity. Although this approach does not overcome all the problems of cross-sectional analysis with endogenous variables, there is ample evidence that it greatly improves inference.

A.6 References

Abowd, J., Creecy, R. & Kramarz, F. (2002), Computing person and firm effects using linked longitudinal employer-employee data, Technical Paper 2002-

06, U.S. Census Bureau, April.

Andrews, M.,T. Schank T., and R. Upward R., (2005), Practical estimation methods for linked employer-employee data", mimeo, May 2005, available from http://www.socialsciences.manchester.ac.uk/disciplines/economics/about/staff/andrews

Albert, P.S., Follmann, D.A., (2000), Modeling repeated count data subject to informative dropout. Biometrics 56, 667–677.

Berndt, E.R., Hall, B.H., Hall, R.E. and Hausmann, J.A., (1974), Estimation and inference in nonlinear structural models, Annals of Economic and Social Measurement, 3, 653-666.

Bartholomew, D.J., (1987), Latent Variable Models and Factor Analysis, Oxford University Press, Oxford.

Bock, R. D., (1985), Contributions of empirical Bayes and marginal maximum likelihood methods to the measurement of individual differences, pages 75-99 of E.E. Roskam (ed), Measurement and Personality Assessment, Elsevier, Amsterdam

Borjas, G.J., Sueyoshi, G.T., (1994), A two-stage estimator for probit models with structural group effects. Journal of Econometrics 64, 165–182.

Breslow, N.E., and Clayton, D., (1993), Approximate inference in generalised linear mixed models. JASA, 88, 9-25.

Browne, W. J., and Draper, D., (2006), A comparison of Bayesian and likelihood-based methods for fitting multilevel models. To appear (with discussion) in Bayesian Analysis

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B., (2003),. Bayesian Data Analysis, 2nd Edition. Chapman and Hall/CRC, Boca Raton, FL.

Lee, L.-F., (2000), A numerically stable quadrature procedure for the one-factor random-component discrete choice model. Journal of Econometrics 95, 117–129.

Lesaffre, E., Spiessens, B., (2001), On the effect of the number of quadrature points in a logistic randomeffects model: an example. Applied Statistics 50, 325–335.

Rabe-Hesketh, S., Skrondal, A. and Pickles, A. (2004), GLLAMM Manual. U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 160.. Downloadable from `http://www.gllamm.org/docum.html`

Rabe-Hesketh, S., Skrondal, A., Pickles, A., (2005), Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects, Journal of Econometrics 128 (2005) 301–323.

Rodriguez, B., and Goldman, N., (1995), An assessment of estimation procedures for multilevel models with binary responses, JRSS, A, 158, 73-89.

Rodriguez, G., and Goldman, N., (2001), Improved estimation procedures for multilevel models with binary response: a case study. Journal of the Royal Statistical Society, A 164, 339–355.

Stroud, A.H., Secrest, D., (1966), Gaussian Quadrature Formulas. Prentice-Hall, Englewood Cliffs, NJ.

Appendix B

Data Preparation for the Analysis of Panel and Clustered Data

Sabre concentrates on procedures for estimating random and fixed effect models, it only has a few commands for performing simple transformations. For instance it does not have the facilities for handling data with missing values, reshaping data, or manipulations like `sort` on a particular variable, so these activities are best performed in more general statistical software packages.

These notes are not meant to be comprehensive as there are many sites that provide an introduction to Stata and some of the commands we present here, see for e.g. `http://www.ats.ucla.edu/stat/stata/`.

B.1 Creation of Dummy Variables

Johnson and Albert (1999) analysed data on the grading of the same essay by five experts. Essays were graded on a scale of 1 to 10 with 10 being excellent. In this exercise we use the subset of the data limited to the grades from graders 1 to 5 on 198 essays (`essays.dta`). The same data were used by Rabe-Hesketh and Skrondal (2005, exercise 5.4).

B.1.1 References

Johnson, V. E., and Albert, J. H., (1999), Ordinal Data Modelling, Springer, New York.

Rabe-Hesketh, S., and Skrondal, A., (2005), Multilevel and Longitudinal Modelling using Stata, Stata Press, Stata Corp, College Station, Texas.

B.1.2 Data description

Number of observations (rows): 990
Number of variables (columns): 11

B.1.3 Variables

`essay`: essay identifier (1,2,...,198}
`grader`: grader identifier {1,2,3,4,5}
`grade`: essay grade {1,2,...,10}
`rating`: essay rate {1,2,...,10}, not used in this exercise
`constant`: 1 for all observations, not used in this exercise
`wordlength`: average word length
`sqrtwords`: square root of the number of words in the essay
`commas`: number of commas times 100 and divided by the number of words in the essay
`errors`: percentage of spelling errors in the essay
`prepos`: percentage of prepositions in the essay
`sentlength`: average length of sentences in the essay

essay	grader	grade	rating	cons	wordlength	sqrtwords	commas	errors	prepos	sentlength
1	1	8	8	1	4.76	15.46	5.60	5.55	8.00	19.53
2	1	7	7	1	4.24	9.06	3.60	1.27	9.50	16.38
3	1	2	2	1	4.09	16.19	1.10	2.61	14.00	18.43
4	1	5	5	1	4.36	7.55	1.80	1.81	0.00	14.65
5	1	7	7	1	4.31	9.64	2.30	0.00	10.00	18.72
6	1	8	10	1	4.51	11.92	1.30	0.00	11.10	20.00
7	1	5	5	1	3.94	8.54	2.80	0.00	13.80	23.75
8	1	2	2	1	4.04	7.21	0.00	0.00	5.90	25.43
9	1	5	5	1	4.24	7.68	5.30	1.72	14.00	28.25
10	1	7	7	1	4.31	8.83	1.30	1.27	14.70	19.28
11	1	5	5	1	4.31	8.77	0.00	1.30	8.00	10.72
12	1	7	7	1	4.69	8.89	3.80	1.31	8.00	13.38
13	1	5	5	1	4.10	8.66	0.00	1.40	5.50	23.71
14	1	6	6	1	4.80	9.69	3.20	7.44	10.90	15.19
15	1	3	3	1	4.06	10.10	1.00	4.08	13.00	24.72
16	1	6	6	1	4.33	13.82	2.10	1.61	11.60	24.05
17	1	5	5	1	4.13	7.55	3.60	0.00	9.00	28.74
18	1	4	4	1	4.07	6.93	2.10	0.00	4.30	15.38
19	1	2	2	1	4.98	6.40	5.20	7.74	12.70	12.74

The first few lines of `essays.dta`

The `essays.dta` dataset contains a variable `grade` which gives the grading of essays on a scale of 1 to 10 (the highest grade given is actually 8 in this data set).

If we want to create a grouping variables/binary indicator/dummy variable for those essays that obtained a `grade` of 5 or over, as compared to those essays that got less than 5 we would use the command

```
pass=1, if grade (5-10), 0 if grade (1-4)
```

we can also do this by using the commands

```
gen pass = 0
replace pass = 1 if grade >= 5
```

The variable `grader` which identifies different examiners and takes the values 1,2,3,4,5. To create dummy variables for examiners 2-5, we can use

```
gen grader2 = 0
replace grader2 = 1 if grader == 2
gen grader3 = 0
replace grader3 = 1 if grader == 3
gen grader4 = 0
replace grader4 = 1 if grader == 4
gen grader5 = 0
replace grader5 = 1 if grader == 5
save essays2, replace
```

essay	grader	grade	rating	constant	wordlength	sqrtwords	commas	errors	prepos	sentlength	pass	grader2	grader3	grader4	grader5
1	3	8	8	1	4.76	15.46	5.60	5.55	8	19.53	1	0	1	0	0
1	1	8	8	1	4.76	15.46	5.60	5.55	8	19.53	1	0	0	0	0
1	4	8	8	1	4.76	15.46	5.60	5.55	8	19.53	1	0	0	1	0
1	2	6	8	1	4.76	15.46	5.60	5.55	8	19.53	1	1	0	0	0
1	5	5	8	1	4.76	15.46	5.60	5.55	8	19.53	1	0	0	0	1
2	2	5	7	1	4.24	9.06	3.60	1.27	9.5	16.38	1	1	0	0	0
2	4	5	7	1	4.24	9.06	3.60	1.27	9.5	16.38	1	0	0	1	0
2	3	3	7	1	4.24	9.06	3.60	1.27	9.5	16.38	0	0	1	0	0
2	1	7	7	1	4.24	9.06	3.60	1.27	9.5	16.38	1	0	0	0	0
2	5	3	7	1	4.24	9.06	3.60	1.27	9.5	16.38	0	0	0	0	1
3	5	1	2	1	4.09	16.19	1.10	2.61	14	18.43	0	0	0	0	1
3	1	2	2	1	4.09	16.19	1.10	2.61	14	18.43	0	0	0	0	0
3	4	1	2	1	4.09	16.19	1.10	2.61	14	18.43	0	0	0	1	0
3	2	1	2	1	4.09	16.19	1.10	2.61	14	18.43	0	1	0	0	0
3	3	1	2	1	4.09	16.19	1.10	2.61	14	18.43	0	0	1	0	0
4	4	5	5	1	4.36	7.55	1.80	1.81	0	14.65	1	0	0	1	0

The first few lines of the new data, `essays2.dta`

This data set can now be read directly into Sabre, see for example, Exercise C3.

B.2 Sorting datasets

Garner and Raudenbush (1991) and Raudenbush and Bryk (2002) studied the role of school and neighbourhood effects on educational attainment. The data set they used (`neighbourhood.dta`) was for young people who left school between 1984 and 1986 from one Scottish Educational Authority. The same data were used by Rabe-Hesketh and Skrondal (2005, exercise 2.2).

B.2.1 References

Garner, C. L., and Raudenbush, S. W., (1991), Neighbourhood effects on educational attainment: A multilevel analysis of the influence of pupil ability, family, school and neighbourhood, Sociology of education, 64, 252-262.

Raudenbush, S. W., and Bryk, A. S., (2002), Hierarchical Linear Models, Sage, Thousand Oaks, CA.

Rabe-Hesketh, S., and Skrondal, A., (2005), Multilevel and Longitudinal Modelling using Stata, Stata Press, Stata Corp, College Station, Texas.

B.2.2 Data description

Number of observations (rows): 2310
Number of variables (columns): 12

B.2.3 Variables:

`neighid`: respondent's neighbourhood identifier
`schid`: respondent's schools identifier
`attain`: respondent's combined end of school educational attainment as measured by grades from various exams
`p7vrq`: respondent's verbal reasoning quotient as measured by a test at age 11-12 in primary school
`p7read`: respondent's reading test score as measured by a test at age 11-12 in primary school
`dadocc`: respondent's father's occupation
`dadunemp`: 1 if respondent's father unemployed, 0 otherwise
`daded`: 1 if respondent's father was in full time education after age 15, 0 otherwise
`momed`: 1 if respondent's mother was in full time education after age 15, 0 otherwise
`male`: 1 if respondent is male, 0 otherwise
`deprive`: index of social deprivation for the local community in which the respondent lived

`dummy`: 1 to 4; representing collections of the schools or neighbourhoods

neighid	schid	attain	p7vrq	p7read	dadocc	dadunemp	daded	momed	male	deprive	dummy
675	0	0.74	21.97	12.13	2.32	0	0	0	1	-0.18	1
647	0	0.26	-7.03	-12.87	16.20	0	0	1	0	0.21	1
650	0	-1.33	-11.03	-31.87	-23.45	1	0	0	1	0.53	1
650	0	0.74	3.97	3.13	2.32	0	0	0	1	0.53	1
648	0	-0.13	-2.03	0.13	-3.45	0	0	0	0	0.19	1
648	0	0.56	-5.03	-0.87	-3.45	0	0	0	0	0.19	1
665	0	-0.36	-2.03	-1.87	16.20	0	0	0	1	0.38	1
661	0	0.74	8.97	3.13	2.32	0	0	0	0	-0.40	1
675	0	-0.36	-2.03	4.13	-3.45	0	1	1	1	-0.18	1
664	0	0.91	16.97	28.13	-3.45	0	0	1	0	-0.17	1
663	0	0.16	-4.03	-8.87	-9.09	0	0	0	1	-0.22	1
661	0	1.52	17.97	25.13	2.32	0	0	0	0	-0.40	1
665	0	0.26	5.97	7.13	-11.49	1	0	0	0	0.38	1
668	0	0.03	0.97	-11.87	2.32	0	0	0	0	-0.24	1
687	0	-0.13	6.97	12.13	-11.49	0	0	0	1	-0.05	1

The first few lines of `neighborood.dta`

The `neighborhood.dta` dataset could be used for random effects models at both the school and neighborhood levels. To obtain separate datasets for each level, with each one sorted on a particular variable (which will be specified as the `case` variable within Sabre) we can use

```
use neighborhood
sort schid
save neighborhood1, replace
sort neighid
save neighborhood2, replace
```

neighid	schid	attain	p7vrq	p7read	dadocc	dadunemp	daded	momed	male	deprive	dummy
26	20	0.56	2.97	6.13	2.32	0	0	0	0	-0.55	4
26	20	1.52	1.97	11.13	-9.09	0	0	0	0	-0.55	4
26	20	-1.33	-1.03	-0.87	-3.45	0	0	0	1	-0.55	4
26	20	1.52	17.97	17.13	16.20	0	0	0	1	-0.55	4
26	20	-1.33	-10.03	-27.87	-3.45	0	0	0	0	-0.55	4
27	20	-0.13	3.97	-0.87	-3.45	0	0	1	0	0.15	4
29	18	0.03	8.97	6.13	16.20	0	1	1	1	-0.08	4
29	20	-0.36	-8.03	-13.87	-3.45	0	0	0	0	-0.08	4
29	20	0.16	4.97	11.13	-11.49	0	0	0	1	-0.08	4
29	20	0.16	1.97	-4.87	-11.49	0	0	0	0	-0.08	4
29	20	0.74	-4.03	0.13	-3.45	0	0	0	1	-0.08	4
29	20	-1.33	-17.03	-23.87	-3.45	0	0	0	1	-0.08	4
29	20	-1.33	-8.03	-4.87	-3.45	0	0	0	0	-0.08	4
29	20	-1.33	-15.03	-25.87	-3.45	0	0	0	0	-0.08	4
29	20	0.56	-0.03	-5.87	-3.45	0	0	0	0	-0.08	4
30	20	-0.13	-0.03	-0.87	-11.49	0	0	0	1	-0.27	4
30	20	1.52	0.97	6.13	29.23	0	1	0	1	-0.27	4
31	18	2.42	2.97	5.13	16.20	0	1	1	1	0.01	4
31	20	-1.33	-16.03	-14.87	-3.45	0	0	0	1	0.01	4
32	20	0.26	0.97	7.13	2.32	0	0	0	1	-0.19	4
32	20	-0.36	-0.03	2.13	-23.45	0	0	0	0	-0.19	4
32	20	-0.36	-1.03	-7.87	-3.45	0	0	0	0	-0.19	4
32	18	0.03	7.97	8.13	-3.45	0	0	0	1	-0.19	4
32	20	0.03	6.97	5.13	-3.45	0	0	0	1	-0.19	4
32	20	0.74	-1.03	7.13	29.23	0	0	0	0	-0.19	4
33	20	-0.36	-4.03	-6.87	-3.45	0	0	0	0	0.54	4

The first few lines of `neighborood2.dta`

This data set can now be read directly into Sabre, see for example, Exercise C2.

B.3 Missing values

Raudenbush and Bhumirat (1992) analysed data on children repeating a grade during their time at primary school. The data were from a national survey of primary education in Thailand in 1988, we use a sub set of that data here.

B.3.1 Reference

Raudenbush, S.W., Bhumirat, C., 1992. The distribution of resources for primary education and its consequences for educational achievement in Thailand, International Journal of Educational Research, 17, 143-164

B.3.2 Data description

Number of observations (rows): 8582
Number of variables (columns): 5

B.3.3 Variables

`schoolid`: school identifier
`sex`: 1 if child is male, 0 otherwise
`pped`: 1 if the child had pre primary experience, 0 otherwise
`repeat`: 1 if the child repeated a grade during primary school, 0 otherwise
`msesc`: mean pupil socio economic status at the school level

schoolid	sex	pped	repeat	msesc
10101	0	1	0	.
10101	0	1	0	.
10101	0	1	0	.
10101	0	1	0	.
10101	0	1	0	.
10101	0	1	0	.
10101	0	1	0	.
10101	0	1	0	.
10101	0	1	0	.
10101	0	1	0	.
10101	0	1	0	.
10101	0	1	0	.
10101	0	1	0	.
10101	0	1	0	.
10101	0	1	0	.
10101	1	1	0	.
10101	1	1	0	.
10101	1	1	0	.
10101	1	1	0	.
10102	0	0	0	.
10102	0	1	0	.
10102	0	1	0	.
10102	0	1	0	.
10102	0	1	0	.
10102	0	1	0	.
10102	0	1	0	.
10102	0	1	0	.
10102	0	1	0	.
10102	0	1	0	.
10102	0	1	0	.
10102	1	1	0	.
10102	1	1	0	.
10102	1	1	0	.
10102	1	1	0	.
10102	1	1	0	.
10102	1	1	0	.
10102	1	1	0	.
10102	1	1	0	.
10102	1	1	0	.
10102	1	1	0	.
10102	1	1	0	.
10102	1	1	0	.
10102	1	1	0	.
10103	0	0	0	0.88
10103	0	0	0	0.88
10103	0	1	0	0.88
10103	0	1	0	0.88

The first few lines of
`thaieduc.dta`

This shows that the `thaieduc.dta` dataset contains a variable `msesc` which has missing values. For models which do not use `msesc` we can simply drop this variable from the dataset as follows

```
use thaieduc
drop msesc
save thaieduc1, replace
```

This dataset has 8,582 observations on 4 variables. For models which do use `msesc` we need to drop all of the missing values. To do this, we can use

```
use thaieduc
drop if msesc ==.
save thaieduc2, replace
```

This dataset has 7,516 observations on 5 variables.

This data set can now be read directly into Sabre, see for example, Example C3.

B.4 Grouped means

The data we use in this example are a sub-sample from the 1982 High School and Beyond Survey (Raudenbush, Bryk, 2002), and include information on 7,185 students nested within 160 schools: 90 public and 70 Catholic. Sample sizes vary from 14 to 67 students per school.

B.4.1 Reference

Raudenbush, S.W., Bryk, A.S., 2002, Heirarchical Linear Models, Thousand Oaks, CA. Sage.

B.4.2 Data description

Number of observations (rows): 7185
Number of variables (columns): 15

B.4.3 Variables

`school`: school identifier
`student`: student identifier
`minority`: 1 if student is from an ethnic minority, 0 otherwise
`gender`: 1 if student is female, 0 otherwise
`ses`: a standardized scale constructed from variables measuring parental education, occupation, and income, socio economic status
`meanses`: mean of the SES values for the students in this school
`mathach`: a measure of the students mathematics achievement
`size`: school enrolment
`sector`: 1 if school is from the Catholic sector, 0 otherwise
`pracad`: proportion of students in the academic track

`disclim`: a scale measuring disciplinary climate
`himnty`:1 if more than 40% minority enrolment, 0 otherwise

school	student	minority	gender	ses	meanses	cses	mathach	size	sector	pracad	disclim	himinty	meansesBYcses	sectorBYcses
1224	1	0	1	-1.53	-0.43	-1.10	5.88	842	0	0.35	1.60	0	0.47	0
1224	2	0	1	-0.59	-0.43	-0.16	19.71	842	0	0.35	1.60	0	0.07	0
1224	3	0	0	-0.53	-0.43	-0.10	20.35	842	0	0.35	1.60	0	0.04	0
1224	4	0	0	-0.67	-0.43	-0.24	8.78	842	0	0.35	1.60	0	0.10	0
1224	5	0	0	-0.16	-0.43	0.27	17.90	842	0	0.35	1.60	0	-0.12	0
1224	6	0	0	0.02	-0.43	0.45	4.58	842	0	0.35	1.60	0	-0.19	0
1224	7	0	1	-0.62	-0.43	-0.19	-2.83	842	0	0.35	1.60	0	0.08	0
1224	8	0	0	-1.00	-0.43	-0.57	0.52	842	0	0.35	1.60	0	0.24	0
1224	9	0	1	-0.89	-0.43	-0.46	1.53	842	0	0.35	1.60	0	0.20	0
1224	10	0	0	-0.46	-0.43	-0.03	21.52	842	0	0.35	1.60	0	0.01	0
1224	11	0	1	-1.45	-0.43	-1.02	9.48	842	0	0.35	1.60	0	0.44	0
1224	12	0	1	-0.66	-0.43	-0.23	16.06	842	0	0.35	1.60	0	0.10	0
1224	13	0	0	-0.47	-0.43	-0.04	21.18	842	0	0.35	1.60	0	0.02	0
1224	14	0	1	-0.99	-0.43	-0.56	20.18	842	0	0.35	1.60	0	0.24	0
1224	15	0	0	0.33	-0.43	0.76	20.35	842	0	0.35	1.60	0	-0.33	0
1224	16	0	1	-0.68	-0.43	-0.25	20.51	842	0	0.35	1.60	0	0.11	0
1224	17	0	0	-0.30	-0.43	0.13	19.34	842	0	0.35	1.60	0	-0.06	0
1224	18	1	0	-1.53	-0.43	-1.10	4.14	842	0	0.35	1.60	0	0.47	0
1224	19	0	1	0.04	-0.43	0.47	2.93	842	0	0.35	1.60	0	-0.20	0
1224	20	0	0	-0.08	-0.43	0.35	16.41	842	0	0.35	1.60	0	-0.15	0
1224	21	0	1	0.06	-0.43	0.49	13.65	842	0	0.35	1.60	0	-0.21	0
1224	22	0	1	-0.13	-0.43	0.30	6.56	842	0	0.35	1.60	0	-0.13	0
1224	23	0	1	0.47	-0.43	0.90	9.65	842	0	0.35	1.60	0	-0.39	0

The first few lines of `hsb.dta`

The `hsb.dta` dataset contains a variable `ses` for each student and the variable `meanses` which is the mean of the SES values for the students in this school. If this school level variable had not been made available with the data set it would need to be created. To create the mean value of 'ses' in Stata for each school based on the students in the sample, we would use the commands

```
sort school
by school:  egen meanses2 = mean(ses)
```

This data set can be used in Sabre, see for example, Examples C1 and C2..

B.5 Reshaping data

Dunn (1992) reported data for the 12-item version of Goldberg's (1972) General Health Questionnaire for psychological distress. The questionnaire was completed by 12 students on 2 dates, 3 days apart. The data are repeated in the table below, the same data were used by Rabe-Hesketh and Skrondal (2005, exercise 1.2).

B.5.1 Data description

Number of observations (rows): 12
Number of variables (columns): 3

B.5.2 Variables

student: student identifier {1,2,...,12}
ghq1: psychological distress score at occasion 1
ghq2: psychological distress score at occasion 2

student	ghq1	ghq2
1	12	12
2	8	7
3	22	24
4	10	14
5	10	8
6	6	4
7	8	5
8	4	6
9	14	14
10	6	5
11	2	5
12	22	16

The data (ghq.dta)

The ghq.dta dataset contains variables ghq1 and ghq2 giving the psychological distress score for students on two separate occasions. To reshape the data from wide into long format and create a single score variable ghq, we can use

```
reshape long ghq, i(student) j(r)
tab r, gen(r)
sort student r
rename r1 dg1
rename r2 dg2
save ghq2, replace
```

This also creates a response indicator variable r, the associated dummy variables dg1 and dg2 and saves the file ghq2.dta.

B.5.3 Variables

r: response occasion 1, 2
student: student identifier {1,2,...,12}
ghq: psychological distress score at occasion
dg1: 1, if the response occasion is 1, 0 otherwise
dg2: 1, if the response occasion is 2, 0 otherwise

The data set was saved as ghq2.dta

r	student	ghq	dg1	dg2
1	1	12	1	0
2	1	12	0	1
1	2	8	1	0
2	2	7	0	1
1	3	22	1	0
2	3	24	0	1
1	4	10	1	0
2	4	14	0	1
1	5	10	1	0
2	5	8	0	1
1	6	6	1	0
2	6	4	0	1
1	7	8	1	0
2	7	5	0	1
1	8	4	1	0
2	8	6	0	1
1	9	14	1	0
2	9	14	0	1
1	10	6	1	0
2	10	5	0	1
1	11	2	1	0
2	11	5	0	1
1	12	22	1	0
2	12	16	0	1

ghq2.dta

This data set can now be read directly into Sabre, see for example Exercise L1.

B.6 Reshaping a data set with a baseline response

In this example we illustrate how to stack a Stata data set (respiratory.dta), which comes with time constant covariates, and a baseline response. We shall suppose that you want to model the baseline jointly with the subsequent responses.

Koch et al (1989) analysed the clinical trial data from 2 centres that compared two groups for respiratory illness. Eligible patients were randomised to treatment or placebo groups at each centre. The respiratory status (ordered response) of each patient prior to randomisation and at 4 later visits to the clinic was determined.

The number of young patients in the sample is 110. The version of the data set (respiratory.dat) we use was also used by Rabe-Hesketh and Skrondal (2005, exercise 5.1).

B.6.1 References

Koch, G. G., Car, G. J., Amara, A., Stokes, M. E., and Uryniak, T. J., (1989), Categorical data analysis. In StateBerry, D., A., Statistical Methodology in the Pharmaceutical Sciences, pp 389-473, Marcel Dekker, New York.

Rabe-Hesketh, S., and Skrondal, A., (2005), Multilevel and Longitudinal Modelling using Stata, Stata Press, Stata Corp, College Station, Texas.

B.6.2 Data description

Number of observations (rows): 110
Number of variables (columns): 10

B.6.3 Variables:

`center`: Centre (1,2)
`drug`: 1 if patient was allocated to the treatment group, 0 otherwise
`male`: 1 if patient was male, 0 otherwise
`age`: patient's age
`bl`: patient's respiratory status prior to randomisation
`v1`: patient's respiratory status at visit 1 (0: terrible; 1: poor; 2: fair; 3: good; 4: excellent)
`v2`: patient's respiratory status at visit 2 (0: terrible; 1: poor; 2: fair; 3: good; 4: excellent)
`v3`: patient's respiratory status at visit 3 (0: terrible; 1: poor; 2: fair; 3: good; 4: excellent)
`v4`: patient's respiratory status at visit 4 (0: terrible; 1: poor; 2: fair; 3: good; 4: excellent)
`patient`: Patient identifier (1,2,...,110)

center	drug	male	age	bl	v1	v2	v3	v4	patient
1	1	0	32	1	2	2	4	2	1
1	1	0	47	2	2	3	4	4	2
1	1	1	11	4	4	4	4	2	3
1	1	1	14	2	3	3	3	2	4
1	1	1	15	0	2	3	3	3	5
1	1	1	20	3	3	2	3	1	6
1	1	1	22	1	2	2	2	3	7
1	1	1	22	2	1	3	4	4	8
1	1	1	23	3	3	4	4	3	9
1	1	1	23	2	3	4	4	4	10
1	1	1	25	2	3	3	2	3	11
1	1	1	26	1	2	2	3	2	12
1	1	1	26	2	2	2	2	2	13
1	1	1	26	2	4	1	4	2	14
1	1	1	28	1	2	2	1	2	15
1	1	1	28	0	0	1	2	1	16
1	1	1	30	3	3	4	4	2	17
1	1	1	30	3	4	4	4	3	18
1	1	1	31	1	2	3	1	1	19
1	1	1	31	3	3	4	4	4	20
1	1	1	31	0	2	3	2	1	21
1	1	1	32	3	4	4	3	3	22
1	1	1	43	1	1	2	1	1	23
1	1	1	46	4	3	4	3	4	24

The first few lines of `respiratory.dta`

The response variables `bl, v1, v2, v3, v4` need to be stacked as a single column

The data is stacked by generating new variables `y1-y5` representing the outcomes (baseline and 4 visits), indexing each observation as `ij` and then reshaping the data from wide to long format. The separate outcomes are thus converted into a single outcome variable `y`. A new variable `r` is also generated which indexes the responses and this can be converted into dummy variables `r1-r5` by use of the '`tab r, gen(r)`' command. The data is sorted by individual, response and observation index. The response indicator is used to generate new baseline and trend covariates from the original baseline measure. Finally, the new data is saved.

We also create a dummy variable `bld=1` if status is from pre-randomisation and a linear trend variable, called `trend =m` if status is from `vm, m=1,2,3,4.` Further we create the variable `base=bl` for each row of post treatment data, 0 for the pre-randomisation data.

B.6.4 Stata Commands

```
use respiratory
gen y1 = bl
gen y2 = v1
gen y3 = v2
gen y4 = v3
gen y5 = v4
reshape long y, i(patient) j(r)
tab r, gen(r)
sort patient r
generate status=y+1
gen bld = r1
gen trend = r-1
gen base = 0
replace base = bl if r >= 2
save respiratory2, replace
```

The command `generate status=y+1` is needed as the ordered response model command in Sabre only work on response variables that use 1 for label of the 1^{st} category.

r	center	drug	male	age	bl	v1	v2	v3	v4	patient	status	r1	r2	r3	r4	r5	bld	trend	base
1	1	1	0	32	1	2	2	4	2	1	2	1	0	0	0	0	1	0	0
2	1	1	0	32	1	2	2	4	2	1	3	0	1	0	0	0	0	1	1
3	1	1	0	32	1	2	2	4	2	1	3	0	0	1	0	0	0	2	1
4	1	1	0	32	1	2	2	4	2	1	5	0	0	0	1	0	0	3	1
5	1	1	0	32	1	2	2	4	2	1	3	0	0	0	0	1	0	4	1
1	1	1	0	47	2	2	3	4	4	2	3	1	0	0	0	0	1	0	0
2	1	1	0	47	2	2	3	4	4	2	3	0	1	0	0	0	0	1	2
3	1	1	0	47	2	2	3	4	4	2	4	0	0	1	0	0	0	2	2
4	1	1	0	47	2	2	3	4	4	2	5	0	0	0	1	0	0	3	2
5	1	1	0	47	2	2	3	4	4	2	5	0	0	0	0	1	0	4	2
1	1	1	1	11	4	4	4	4	2	3	5	1	0	0	0	0	1	0	0
2	1	1	1	11	4	4	4	4	2	3	5	0	1	0	0	0	0	1	4
3	1	1	1	11	4	4	4	4	2	3	5	0	0	1	0	0	0	2	4
4	1	1	1	11	4	4	4	4	2	3	5	0	0	0	1	0	0	3	4
5	1	1	1	11	4	4	4	4	2	3	3	0	0	0	0	1	0	4	4
1	1	1	1	14	2	3	3	3	2	4	3	1	0	0	0	0	1	0	0
2	1	1	1	14	2	3	3	3	2	4	4	0	1	0	0	0	0	1	2
3	1	1	1	14	2	3	3	3	2	4	4	0	0	1	0	0	0	2	2
4	1	1	1	14	2	3	3	3	2	4	4	0	0	0	1	0	0	3	2
5	1	1	1	14	2	3	3	3	2	4	3	0	0	0	0	1	0	4	2
1	1	1	1	15	0	2	3	3	3	5	1	1	0	0	0	0	1	0	0
2	1	1	1	15	0	2	3	3	3	5	3	0	1	0	0	0	0	1	0
3	1	1	1	15	0	2	3	3	3	5	4	0	0	1	0	0	0	2	0
4	1	1	1	15	0	2	3	3	3	5	4	0	0	0	1	0	0	3	0
5	1	1	1	15	0	2	3	3	3	5	4	0	0	0	0	1	0	4	0
1	1	1	1	20	3	3	2	3	1	6	4	1	0	0	0	0	1	0	0

The first few lines of the resulting data set, `respiratory2.dta`

This data set can now be read directly into Sabre, see for example Exercise L6.

B.7 Creating Data Sets for Bivariate Models

Vella and Verbeek (1998) analysed the male data from the Youth Sample of the US National Longitudinal Survey for the period 1980-1987. The number of young males in the sample is 545. The version of the data set (wagepan.dta) we use was obtained from Wooldridge (2002). The same data were used for modelling the wages and for separately modelling trade union membership by Rabe-Hesketh and Skrondal (2005, exercises 2.7 and 4.7). We start by looking at the data set (`wagepan.dta`) in a form appropriate for response panel models of log(wage) or union membership.

B.7.1 Data description

Number of observations (rows): 4360
Number of variables (columns): 44

B.7.2 Variables

`nr`: person identifier
`year`: 1980 to 1987
`black`: 1 if respondent is black, 0 otherwise
`exper`: labour market experience (age-6-educ)
`hisp`: 1 if respondent is Hispanic, 0 otherwise
`poorhlth`: 1 if respondent has a health disability, 0 otherwise
`married`: 1 if respondent is married, 0 otherwise
`nrthcen`: 1 if respondent lives in the Northern Central part of the US, 0 otherwise
`nrtheast`: 1 if respondent lives in the North East part of the US, 0 otherwise
`rur`: 1 if respondent lives in a rural area, 0 otherwise
`south`: 1 if respondent lives in the South of the US, 0 otherwise
`educ`: years of schooling
`union`: 1 if the respondent is a member of a trade union, 0 otherwise
`lwage`: log of hourly wage in US dollars
`d8m`:1 if the year is 198m, 0 otherwise, m=1,. . . ,7

nr	year	agric	black	bus	construc	ent	exper	fin	hisp
13	1980	0	0	1	0	0	1	0	0
13	1981	0	0	0	0	0	2	0	0
13	1982	0	0	1	0	0	3	0	0
13	1983	0	0	1	0	0	4	0	0
13	1984	0	0	0	0	0	5	0	0
13	1985	0	0	1	0	0	6	0	0
13	1986	0	0	1	0	0	7	0	0
13	1987	0	0	1	0	0	8	0	0
17	1980	0	0	0	0	0	4	0	0
17	1981	0	0	0	0	0	5	0	0
17	1982	0	0	0	0	0	6	0	0
17	1983	0	0	0	0	0	7	0	0
17	1984	0	0	0	0	0	8	0	0
17	1985	0	0	0	1	0	9	0	0
17	1986	0	0	0	1	0	10	0	0
17	1987	0	0	0	1	0	11	0	0
18	1980	0	0	0	0	0	4	0	0
18	1981	0	0	0	0	0	5	0	0
18	1982	0	0	0	0	0	6	0	0
18	1983	0	0	0	0	0	7	0	0
18	1984	0	0	0	0	0	8	0	0

The first few rows and columns of wagepan.dta

The `wagepan.dta` dataset contains response variables `union` and `lwage` for use in univariate models of union membership and log wages respectively. To create a dataset suitable for a bivariate model of union membership and log wages, we can use

```
gen y1 = union
gen y2 = lwage
gen ij = _n
reshape long y, i(ij) j(r)
tab r, gen(r)
sort nr r ij
save union-wage, replace
```

This creates a single response variable `y`, a response indicator `r` and associated dummy variables `r1` and `r2`. The dataset is sorted by individual `nr`, response indicator `r` and observation index `ij`.

ij	r	nr	year	agric	black	bus	construc	ent	exper	fin	hisp	poorhlth	hours	manuf	married	min	nrthcen	nrtheast	occ1	occ2	occ3	occ4	occ5
1	1	13	1980	0	0	1	0	0	1	0	0	0	2672	0	0	0	0	1	0	0	0	0	0
2	1	13	1981	0	0	0	0	0	2	0	0	0	2320	0	0	0	0	1	0	0	0	0	0
3	1	13	1982	0	0	1	0	0	3	0	0	0	2940	0	0	0	0	1	0	0	0	0	0
4	1	13	1983	0	0	1	0	0	4	0	0	0	2960	0	0	0	0	1	0	0	0	0	0
5	1	13	1984	0	0	0	0	0	5	0	0	0	3071	0	0	0	0	1	0	0	0	0	1
6	1	13	1985	0	0	1	0	0	6	0	0	0	2864	0	0	0	0	1	0	1	0	0	0
7	1	13	1986	0	0	1	0	0	7	0	0	0	2994	0	0	0	0	1	0	1	0	0	0
8	1	13	1987	0	0	1	0	0	8	0	0	0	2640	0	0	0	0	1	0	1	0	0	0
1	2	13	1980	0	0	1	0	0	1	0	0	0	2672	0	0	0	0	1	0	0	0	0	0
2	2	13	1981	0	0	0	0	0	2	0	0	0	2320	0	0	0	0	1	0	0	0	0	0
3	2	13	1982	0	0	1	0	0	3	0	0	0	2940	0	0	0	0	1	0	0	0	0	0
4	2	13	1983	0	0	1	0	0	4	0	0	0	2960	0	0	0	0	1	0	0	0	0	0
5	2	13	1984	0	0	0	0	0	5	0	0	0	3071	0	0	0	0	1	0	0	0	0	1
6	2	13	1985	0	0	1	0	0	6	0	0	0	2864	0	0	0	0	1	0	1	0	0	0
7	2	13	1986	0	0	1	0	0	7	0	0	0	2994	0	0	0	0	1	0	1	0	0	0
8	2	13	1987	0	0	1	0	0	8	0	0	0	2640	0	0	0	0	1	0	1	0	0	0
9	1	17	1980	0	0	0	0	0	4	0	0	0	2484	0	0	0	0	1	0	1	0	0	0
10	1	17	1981	0	0	0	0	0	5	0	0	0	2804	0	0	0	0	1	0	1	0	0	0
11	1	17	1982	0	0	0	0	0	6	0	0	0	2530	0	0	0	0	1	0	1	0	0	0
12	1	17	1983	0	0	0	0	0	7	0	0	0	2340	0	0	0	0	1	0	1	0	0	0

The first lines and columns of `union-wage.dta`

This data set is used in Exercise L10.