6

Multilevel Models for Ordinal and Nominal Variables

Donald Hedeker

University of Illinois at Chicago

6.1 Introduction

Reflecting the usefulness of multilevel analysis and the importance of categorical outcomes in many areas of research, generalization of multilevel models for categorical outcomes has been an active area of statistical research. For dichotomous response data, several approaches adopting either a logistic or probit regression model and various methods for incorporating and estimating the influence of the random effects have been developed [9, 21, 34, 37, 103, 115]. Several review articles [31, 39, 76, 90] have discussed and compared some of these models and their estimation procedures. Also, Snijders and Bosker [99, chap. 14] provide a practical summary of the multilevel logistic regression model and the various procedures for estimating its parameters. As these sources indicate, the multilevel logistic regression model is a very popular choice for analysis of dichotomous data.

Extending the methods for dichotomous responses to ordinal response data has also been actively pursued [4, 29, 30, 44, 48, 58, 106, 113]. Again, developments have been mainly in terms of logistic and probit regression models, and many of these are reviewed in Agresti and Natarajan [5]. Because the proportional odds model described by McCullagh [71], which is based on the logistic regression formulation, is a common choice for analysis of ordinal data, many of the multilevel models for ordinal data are generalizations of this model. The proportional odds model characterizes the ordinal responses in C categories in terms of $C-1$ cumulative category comparisons, specifically, $C-1$ cumulative logits (i.e., log odds) of the ordinal responses. In the proportional odds model, the covariate effects are assumed to be the same across these cumulative logits, or proportional across the cumulative odds. As noted by Peterson and Harrell [77], however, examples of non-proportional odds are

Handbook of Multilevel Analysis, edited by Jan de Leeuw and Erik Meijer

not difficult to find. To overcome this limitation, Hedeker and Mermelstein [52] described an extension of the multilevel ordinal logistic regression model to allow for non-proportional odds for a set of regressors.

For nominal responses, there have been developments in terms of multilevel models as well. An early example is the model for nominal educational test data described by Bock [14]. This model includes a random effect for the level-2 subjects and fixed item parameters for the level-1 item responses nested within subjects. While Bock's model is a full-information maximum likelihood approach, using Gauss-Hermite quadrature to integrate over the random-effects distribution, it doesn't include covariates or multiple random effects. As a result, its usefulness for multilevel modeling is very limited. More general regression models of multilevel nominal data have been considered by Daniels and Gatsonis [25], Revelt and Train [88], Bhat [13], Skrondal and Rabe-Hesketh [97], and in Goldstein [38, chap. 4]. In these models, it is common to adopt a reference cell approach in which one of the categories is chosen as the reference cell and parameters are characterized in terms of the remaining $C-1$ comparisons to this reference cell. Alternatively, Hedeker [47] adopts the approach in Bock's model, which allows any set of $C-1$ comparisons across the nominal response categories. Hartzel et al. [43] synthesizes some of the work in this area, describing a general mixed-effects model for both clustered ordinal and nominal responses, and Agresti et al. [3] describe a variety of social science applications of multilevel modeling of categorical responses.

This chapter describes multilevel models for categorical data that accommodate multiple random effects and allow for a general form for model covariates. Although only 2-level models will be considered here, 3-level generalizations are possible [35, 63, 83, 107]. For ordinal outcomes, proportional odds, partial proportional odds, and related survival analysis models for discrete or grouped-time survival data are described. For nominal response data, models using both reference cell and more general category comparisons are described. Connections with item response theory (IRT) models are also made. A full maximum likelihood solution is outlined for parameter estimation. In this solution, multi-dimensional quadrature is used to numerically integrate over the distribution of random-effects, and an iterative Fisher scoring algorithm is used to solve the likelihood equations. To illustrate application of the various multilevel models for categorical responses, several analyses of a longitudinal psychiatric dataset are described.

6.2 Multilevel Logistic Regression Model

Before considering models for ordinal and nominal responses, the multilevel model for dichotomous responses will be described. This is useful because both the ordinal and nominal models can be viewed as different ways of generalizing the dichotomous response model. To set the notation, let j denote the level-2 units (clusters) and let i denote the level-1 units (nested observations). Assume that there are $j = 1, \ldots, N$ level-2 units and $i = 1, \ldots, n_j$ level-1 units nested within each level-2 unit. The total number of level-1 observations across level-2 units is given by $n = \sum_{j=1}^{N} n_j$. Let $\underline{Y}_{ij}$ be the value of the dichotomous outcome variable, coded 0 or 1, associated with level-1 unit i nested within level-2 unit j. The logistic regression model is written in terms of the log odds (i.e., the logit) of the probability of a response, denoted $\underline{p}_{ij} = \Pr(\underline{Y}_{ij} = 1)$. Augmenting the standard logistic regression model with a single random effect yields

$$\log\left[\frac{\underline{p}_{ij}}{1-\underline{p}_{ij}}\right] = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \underline{\delta}_j ,$$

where $\boldsymbol{x}_{ij}$ is the $s \times 1$ covariate vector (includes a 1 for the intercept), $\boldsymbol{\beta}$ is the $s \times 1$ vector of unknown regression parameters, and $\underline{\delta}_j$ is the random cluster effect (one for each level-2 cluster). These are assumed to be distributed in the population as $\mathcal{N}(0, \sigma_\delta^2)$. For convenience and computational simplicity, in models for categorical outcomes the random effects are typically expressed in standardized form. For this, $\underline{\delta}_j = \sigma_\delta \, \underline{\theta}_j$ and the model is given as

$$\log\left[\frac{\underline{p}_{ij}}{1-\underline{p}_{ij}}\right] = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \sigma_\delta \, \underline{\theta}_j .$$

Notice that the random-effects variance term (i.e., the population standard deviation σ_δ) is now explicitly included in the regression model. Thus, it and the regression coefficients are on the same scale, namely, in terms of the log-odds of a response.

The model can be easily extended to include multiple random effects. For this, denote $\boldsymbol{z}_{ij}$ as the $r \times 1$ vector of random-effect variables (a column of ones is usually included for the random intercept). The vector of random effects $\underline{\boldsymbol{\delta}}_j$ is assumed to follow a multivariate normal distribution with mean vector $\boldsymbol{\emptyset}$ and variance-covariance matrix $\boldsymbol{\Omega}$. To standardize the multiple random effects $\underline{\boldsymbol{\delta}}_j = \boldsymbol{T}\underline{\boldsymbol{\theta}}_j$, where $\boldsymbol{T}\boldsymbol{T}' = \boldsymbol{\Omega}$ is the Cholesky decomposition of $\boldsymbol{\Omega}$. The model is now written as

$$\log\left[\frac{\underline{p}_{ij}}{1-\underline{p}_{ij}}\right] = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{T}\underline{\boldsymbol{\theta}}_j . \tag{6.1}$$

As a result of the transformation, the Cholesky factor $\boldsymbol{T}$ is usually estimated instead of the variance-covariance matrix $\boldsymbol{\Omega}$. As the Cholesky factor is essentially the matrix square-root of the variance-covariance matrix, this allows more stable estimation of near-zero variance terms.

6.2.1 Threshold Concept

Dichotomous regression models are often motivated and described using the "threshold concept" [15]. This is also termed a latent variable model for dichotomous variables [65]. For this, it is assumed that a continuous latent variable $\underline{y}$ underlies the observed dichotomous response $\underline{Y}$. A threshold, denoted γ, then determines if the dichotomous response $\underline{Y}$ equals 0 ($\underline{y}_{ij} \leq \gamma$) or 1 ($\underline{y}_{ij} > \gamma$). Without loss of generality, it is common to fix the location of the underlying latent variable by setting the threshold equal to zero (i.e., $\gamma = 0$). Figure 6.1 illustrates this concept assuming that the continuous latent variable $\underline{y}$ follows either a normal or logistic probability density function (pdf).

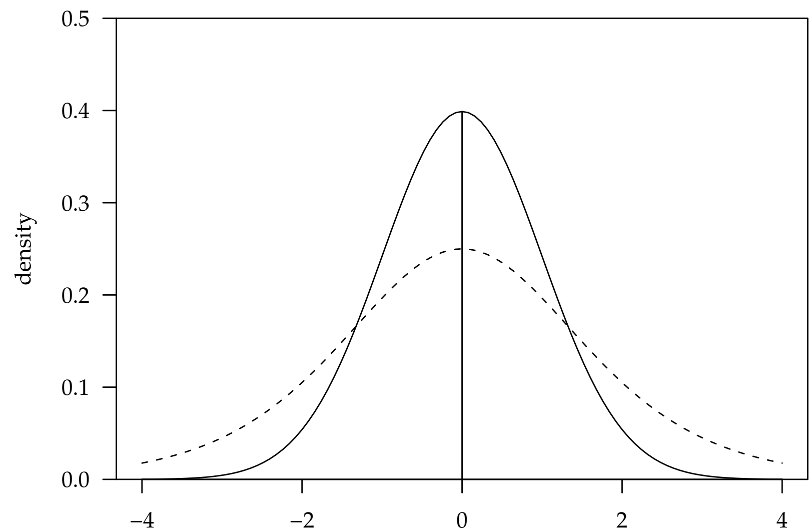

Fig. 6.1. Threshold concept for a dichotomous response (solid = normal, dashed = logistic).

As noted by McCullagh and Nelder [72], the assumption of a continuous latent distribution, while providing a useful motivating concept, is not a strict model requirement. In terms of the continuous latent variable $\underline{y}$, the model is written as

$$\underline{y}_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{T}\underline{\boldsymbol{\theta}}_j + \underline{\epsilon}_{ij} \,.$$

Note the inclusion of the errors $\underline{\epsilon}_{ij}$ in this representation of the model. In the logistic regression formulation, the errors $\underline{\epsilon}_{ij}$ are assumed to follow a standard logistic distribution with mean 0 and variance $\pi^2/3$ [2, 65]. The scale of the errors is fixed because $\underline{y}$ is not observed, and so the the scale is not separately identified. Thus, although the above model appears to be the same as an ordinary multilevel regression model for continuous outcomes, it is one in which the error variance is fixed and not estimated. This has certain consequences that will be discussed later.

Because the errors are assumed to follow a logistic distribution and the random effects a normal distribution, this model and models closely related to it are often referred to as logistic/normal or logit/normit models, especially in the latent trait model literature [11]. If the errors are assumed to follow a normal distribution, then the resulting model is a multilevel probit regression or normal/normal model. In the probit model, the errors have mean 0 and variance 1 (i.e., the variance of the standard normal distribution).

6.2.2 Multilevel Representation

For a multilevel representation of a simple model with only one level-1 covariate x_{ij} and one level-2 covariate x_j, the level-1 model is written in terms of the logit as

$$\log\left[\frac{\underline{p}_{ij}}{1-\underline{p}_{ij}}\right] = \underline{\beta}_{0j} + \underline{\beta}_{1j} x_{ij} \, ,$$

or in terms of the latent response variable as

$$\underline{y}_{ij} = \underline{\beta}_{0j} + \underline{\beta}_{1j} x_{ij} + \underline{\epsilon}_{ij} \, . \tag{6.2}$$

The level-2 model is then (assuming x_{ij} is a random-effects variable)

$$\underline{\beta}_{0j} = \beta_0 + \beta_2 x_j + \underline{\delta}_{0j} \, , \tag{6.3a}$$

$$\underline{\beta}_{1j} = \beta_1 + \beta_3 x_j + \underline{\delta}_{1j} \, . \tag{6.3b}$$

Notice that it's easiest, and in agreement with the normal-theory (continuous) multilevel model, to write the level-2 model in terms of the unstandardized random effects, which are distributed in the population as $\underline{\boldsymbol{\delta}}_j \sim \mathcal{N}(\boldsymbol{\emptyset}, \boldsymbol{\Omega})$. For models with multiple variables at either level-1 or level-2, the above level-1 and level-2 submodels are generalized in an obvious way.

Because the level-1 variance is fixed, the model operates somewhat differently than the more standard normal-theory multilevel model for continuous outcomes. For example, in an ordinary multilevel model the level-1 variance term is typically reduced as level-1 covariates x_{ij} are added to the model. However, this cannot happen in the above model because the level-1 variance is fixed. As noted by Snijders and Bosker [99], what happens instead (as level-1 covariates are added) is that the random-effect variance terms tend to become larger as do the other regression coefficients, the latter become larger in absolute value.

6.2.3 Logistic and Probit Response Functions

The logistic model can also be written as

$$\underline{p}_{ij} = \Psi(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{T}\underline{\boldsymbol{\theta}}_j)\,,$$

where $\Psi(\eta)$ is the logistic cumulative distribution function (cdf), namely

$$\Psi(\eta) = \frac{\exp(\eta)}{1+\exp(\eta)} = \frac{1}{1+\exp(-\eta)}\,.$$

The cdf is also termed the response function of the model. A mathematical nicety of the logistic distribution is that the probability density function (pdf) is related to the cdf in a simple way, namely, $\psi(\eta) = \Psi(\eta)[1-\Psi(\eta)]$.

As mentioned, the probit model, which is based on the standard normal distribution, is often proposed as an alternative to the logistic model. For the probit model, the normal cdf $\Phi(\eta)$ and pdf $\phi(\eta)$ replace their logistic counterparts, and because the standard normal distribution has variance equal to one, $\underline{\epsilon}_{ij} \sim \mathcal{N}(0,1)$. As a result, in the probit model the underlying latent variable vector $\underline{\boldsymbol{y}}_j$ is distributed normally in the population with mean $\boldsymbol{X}_j\boldsymbol{\beta}$ and variance covariance matrix $\boldsymbol{Z}_j\boldsymbol{T}\boldsymbol{T}'\boldsymbol{Z}'_j + \boldsymbol{I}$. The latter, when converted to a correlation matrix, yields tetrachoric correlations for the underlying latent variable vector $\underline{\boldsymbol{y}}$ (and polychoric correlations for ordinal outcomes, discussed below). For this reason, in some areas, for example familial studies, the probit formulation is preferred to its logistic counterpart.

As can be seen in the earlier figure, both the logistic and normal distributions are symmetric around zero and differ primarily in terms of their scale; the standard normal has standard deviation equal to 1, whereas the standard logistic has standard deviation equal to $\pi/\sqrt{3}$. As a result, the two typically give very similar results and conclusions, though the logistic regression parameters (and associated standard errors) are approximately $\pi/\sqrt{3}$ times as large because of the scale difference between the two distributions. An alternative response function, that provides connections with proportional hazards survival analysis models (see Allison [7] and section 6.3.2), is the complementary log-log response function $1 - \exp[-\exp(\eta)]$. Unlike the logistic and normal, the distribution that underlies the complementary log-log response function is asymmetric and has variance equal to $\pi^2/6$. Its pdf is given by $\exp(\eta)[1-p(\eta)]$. As Doksum and Gasko [26] note, large amounts of high quality data are often necessary for response function selection to be relevant. Since these response functions often provide similar fits and conclusions, McCullagh [71] suggests that response function choice should be based primarily on ease of interpretation.

6.3 Multilevel Proportional Odds Model

Let the C ordered response categories be coded as $c = 1, 2, \ldots, C$. Ordinal response models often utilize cumulative comparisons of the ordinal outcome. The cumulative probabilities for the C categories of the ordinal outcome $\underline{Y}$ are defined as $\underline{P}_{ijc} = \Pr(\underline{Y}_{ij} \leq c) = \sum_{k=1}^{c} \underline{p}_{ijk}$. The multilevel logistic model for the cumulative probabilities is given in terms of the cumulative logits as

$$\log\left[\frac{\underline{P}_{ijc}}{1-\underline{P}_{ijc}}\right] = \gamma_c - \left[\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{T}\underline{\boldsymbol{\theta}}_j\right] \qquad (c = 1, \ldots, C-1), \qquad (6.4)$$

with $C-1$ strictly increasing model thresholds γ_c (i.e., $\gamma_1 < \gamma_2 \ldots < \gamma_{C-1}$).

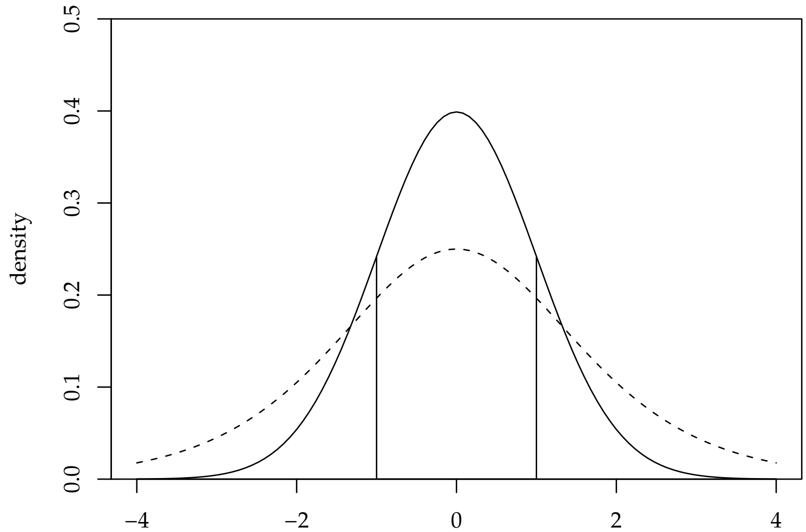

Fig. 6.2. Threshold concept for an ordinal response with 3 categories (solid = normal, dashed = logistic).

The relationship between the latent continuous variable $\underline{y}$ and an ordinal outcome with three categories is depicted in Figure 6.2. In this case, the ordinal outcome $\underline{Y}_{ij} = c$ if $\gamma_{c-1} \leq \underline{y}_{ij} < \gamma_c$ for the latent variable (with $\gamma_0 = -\infty$ and $\gamma_C = \infty$). As in the dichotomous case, it is common to set a threshold to zero to set the location of the latent variable. Typically, this is done in terms of the first threshold (i.e., $\gamma_1 = 0$). In Figure 6.2, setting $\gamma_1 = 0$ implies that $\gamma_2 = 2$.

At first glance, it may appear that the parameterization of the model in (6.4) is not consistent with the dichotomous model in (6.1). To see the connection, notice that for a dichotomous outcome (coded 0 and 1), the model is written as

$$\log\left[\frac{\underline{P}_{ij0}}{1-\underline{P}_{ij0}}\right] = 0 - \left[\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{T}\underline{\boldsymbol{\theta}}_j\right],$$

and since for a dichotomous outcome $\underline{P}_{ij0} = \underline{p}_{ij0}$ and $1 - \underline{P}_{ij0} = \underline{p}_{ij1}$,

$$\log\left[\frac{1-\underline{P}_{ij0}}{\underline{P}_{ij0}}\right] = \log\left[\frac{\underline{p}_{ij1}}{1-\underline{p}_{ij1}}\right] = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{T}\underline{\boldsymbol{\theta}}_j,$$

which is the same as before. Also, in terms of the underlying latent variable $\underline{y}$, the multilevel representation of the ordinal model is identical to the dichotomous version presented earlier in equation (6.2). If the multilevel model is written in terms of the observed response variable $\underline{Y}$, then the level-1 model is written instead as

$$\log\left[\frac{\underline{P}_{ijc}}{1-\underline{P}_{ijc}}\right] = \gamma_c - \left[\underline{\beta}_{0j} + \underline{\beta}_{1j}x_{ij}\right],$$

for the case of a model with one level-1 covariate. Because the level-2 model does not really depend on the response function or variable, it would be the same as given above for the dichotomous model in equations (6.3a) and (6.3b).

Since the regression coefficients $\boldsymbol{\beta}$ 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. McCullagh [71] calls this assumption of identical odds ratios across the $C-1$ cut-offs the proportional odds assumption. 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. 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 $\underline{Y}$, it is not the only way of writing the model. In particular, the model is sometimes written as

$$\log\left[\frac{\underline{P}_{ijc}}{1-\underline{P}_{ijc}}\right] = \gamma_c + \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{T}\underline{\boldsymbol{\theta}}_j \qquad (c = 1, \ldots, C-1),$$

in which case the regression parameters $\boldsymbol{\beta}$ are identical but of opposite sign. This alternate specification is commonly used in survival analysis models (see section 6.3.2).

6.3.1 Partial Proportional Odds

As noted by Peterson and Harrell [77], violation of the proportional odds assumption is not uncommon. Thus, they described a (fixed-effects) partial proportional odds model in which covariates are allowed to have differential effects on the $C-1$ cumulative logits. Similarly, Terza [109] developed a similar extension of the (fixed-effects) ordinal probit model. Hedeker and Mermelstein [52, 53] utilize this extension within the context of a multilevel ordinal regression model. For this, the model for the $C-1$ cumulative logits can be written as

$$\log\left[\frac{\underline{P}_{ijc}}{1-\underline{P}_{ijc}}\right] = \gamma_c - \left[(\boldsymbol{x}^*_{ij})'\boldsymbol{\beta}_c + \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{T}\underline{\boldsymbol{\theta}}_i\right] \qquad (c = 1,\ldots,C-1),$$

where $\boldsymbol{x}^*_{ij}$ is a $h \times 1$ vector containing the values of observation ij on the set of h covariates for which proportional odds is not assumed. In this model, $\boldsymbol{\beta}_c$ is a $h \times 1$ vector of regression coefficients associated with these h covariates. Because $\boldsymbol{\beta}_c$ carries the c subscript, the effects of these h covariates are allowed to vary across the $C-1$ cumulative logits. In many areas of research, this extended model is useful. For example, suppose that in a alchohol reduction study there are three response categories (abstinence, mild use, heavy use) and suppose that an intervention designed to reduce drinking is not successful in increasing the proportion of individuals in the abstinence category but is successful in moving individuals from heavy to mild use. In this case, the (covariate) effect of intervention group would not be observed on the first cumulative logit, but would be observed on the second cumulative logit. This extended model has been utilized in several articles [32, 114, 117], and a similar Bayesian hierarchical model is described in Ishwaran [57].

In general, this extension of the proportional odds model is not problematic, however, one caveat should be mentioned. For the explanatory variables without proportional odds, the effects on the cumulative log odds, namely $(\boldsymbol{x}^*_{ij})'\boldsymbol{\beta}_c$, result in $C-1$ non-parallel regression lines. These regression lines inevitably cross for some values of $\boldsymbol{x}^*$, leading to negative fitted values for the response probabilities. For $\boldsymbol{x}^*$ variables contrasting two levels of an explanatory variable (e.g., gender coded as 0 or 1), this crossing of regression lines occurs outside the range of admissible values (i.e., < 0 or > 1). However, if the explanatory variable is continuous, this crossing can occur within the range of the data, and so, allowing for non-proportional odds can be problematic. A solution to this dilemma is sometimes possible if the variable has, say, m levels with a reasonable number of observations at each of these m levels. In this case $m-1$ dummy-coded variables can be created and substituted into the model in place of the continuous variable. Alternatively, one might consider a nominal response model using Helmert contrasts [15] for the outcome variable. This approach, described in section 6.4, is akin to the sequential logit models for nested or hierarchical response scales described in McCullagh and Nelder [72].

6.3.2 Survival Analysis Models

Several authors have noted the connection between survival analysis models and binary and ordinal regression models for survival data that are discrete or grouped within time intervals (for practical introductions see Allison [6, 7], D'Agostino et al. [24], Singer and Willett [95]). This connection has been utilized in the context of categorical multilevel or mixed-effects regression models by many authors as well [42, 54, 94, 106, 108]. For this, assume that time (of assessment) can take on only discrete positive values $c = 1, 2, \ldots, C$.[1] For each level-1 unit, observation continues until time $\underline{Y}_{ij}$ at which point either an event occurs ($\underline{d}_{ij} = 1$) or the observation is censored ($\underline{d}_{ij} = 0$), where censoring indicates being observed at c but not at $c+1$. Define P_{ijc} to be the probability of failure, up to and including time interval c, that is,

$$P_{ijc} = \Pr(\underline{Y}_{ij} \leq c),$$

and so the probability of survival beyond time interval c is simply $1 - P_{ijc}$.

Because $1 - P_{ijc}$ represents the survivor function, McCullagh [71] proposed the following grouped-time version of the continuous-time proportional hazards model

$$\log[-\log(1 - P_{ijc})] = \gamma_c + \boldsymbol{x}'_{ij}\boldsymbol{\beta}. \tag{6.5}$$

This is the aforementioned complementary log-log response function, which can be re-expressed in terms of the cumulative failure probability, $P_{ijc} = 1 - \exp(-\exp(\gamma_c + \boldsymbol{x}'_{ij}\boldsymbol{\beta}))$. In this model, $\boldsymbol{x}_{ij}$ includes covariates that vary either at level 1 or 2, however they do not vary with time (i.e., they do not vary across the ordered response categories). They may, however, represent the average of a variable across time or the value of the covariate at the time of the event.

The covariate effects in this model are identical to those in the grouped-time version of the proportional hazards model described by Prentice and Gloeckler [79]. As such, the $\boldsymbol{\beta}$ coefficients are also identical to the coefficients in the underlying continuous-time proportional hazards model. Furthermore, as noted by Allison [6], the regression coefficients of the model are invariant to interval length. Augmenting the coefficients $\boldsymbol{\beta}$, the threshold terms γ_c represent the logarithm of the integrated baseline hazard (i.e., when $\boldsymbol{x} = \boldsymbol{\emptyset}$). While the above model is the same as that described in McCullagh [71], it is written so that the covariate effects are of the same sign as the Cox proportional hazards model. A positive coefficient for a regressor then reflects increasing hazard (i.e., lower values of $\underline{Y}$) with greater values of the regressor. Adding (standardized) random effects, we get

---

[1] To make the connection to ordinal models more direct, time is here denoted as c, however more commonly it is denoted as t in the survival analysis literature.

$$\log[-\log(1-\underline{P}_{ijc})] = \gamma_c + \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{T}\underline{\boldsymbol{\theta}}_j\,. \tag{6.6}$$

This model is thus a multilevel ordinal regression model with a complementary log-log response function instead of the logistic. Though the logistic model has also been proposed for analysis of grouped and/or discrete time survival data, its regression coefficients are not invariant to time interval length and it requires the intervals to be of equal length [6]. As a result, the complementary log-log response function is generally preferred.

In the ordinal treatment, survival time is represented by the ordered outcome $\underline{Y}_{ij}$, which is designated as being censored or not. Alternatively, each survival time can be represented as a set of dichotomous dummy codes indicating whether or not the observation failed in each time interval that was experienced [6, 24, 95]. Specifically, each survival time $\underline{Y}_{ij}$ is represented as a vector with all zeros except for its last element, which is equal to $\underline{d}_{ij}$ (i.e., $= 0$ if censored and $= 1$ for an event). The length of the vector for observation ij equals the observed value of $\underline{Y}_{ij}$ (assuming that the survival times are coded as $1, 2, \ldots, C$). These multiple time indicators are then treated as distinct observations in a dichotomous regression model. In a multilevel model, a given cluster's response vector $\underline{\boldsymbol{Y}}_j$ is then of size $(\sum_{i=1}^{n_j} \underline{Y}_{ij}) \times 1$. This method has been called the pooling of repeated observations method by Cupples et al. [23]. It is particularly useful for handling time-dependent covariates and fitting non-proportional hazards models because the covariate values can change across time. See Singer and Willett [96] for a detailed treatment of this method.

For this dichotomous approach, define $\underline{\lambda}_{ijc}$ to be the probability of failure in time interval c, conditional on survival prior to c,

$$\underline{\lambda}_{ijc} = \Pr(\underline{Y}_{ij} = c \mid \underline{Y}_{ij} \geq c).$$

Similarly, $1 - \underline{\lambda}_{ijc}$ is the probability of survival beyond time interval c, conditional on survival prior to c. The multilevel proportional hazards model is then written as

$$\log[-\log(1-\underline{\lambda}_{ijc})] = \boldsymbol{x}'_{ijc}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{T}\underline{\boldsymbol{\theta}}_j\,, \tag{6.7}$$

where now the covariates $\boldsymbol{x}$ can vary across time and so are denoted as $\boldsymbol{x}_{ijc}$. The first elements of $\boldsymbol{x}$ are usually timepoint dummy codes. Because the covariate vector $\boldsymbol{x}$ now varies with c, this approach automatically allows for time-dependent covariates, and relaxing the proportional hazards assumption only involves including interactions of covariates with the timepoint dummy codes.

Under the complementary log-log link function, the two approaches characterized by (6.6) and (6.7) yield identical results for the parameters that do not depend on c [28, 59]. Comparing these two approaches, notice that for the ordinal approach each observation consists of only two pieces of data: the (ordinal) time of the event and whether it was censored or not. Alternatively, in the dichotomous approach each survival time is represented as a vector of dichotomous indicators, where the size of the vector depends upon the timing of the event (or censoring). Thus, the ordinal approach can be easier to implement and offers savings in terms of the dataset size, especially as the number of timepoints gets large, while the dichotomous approach is superior in its treatment of time-dependent covariates and relaxing of the proportional hazards assumption.

6.3.3 Estimation

For the ordinal models presented, the probability of a response in category c for a given level-2 unit j, conditional on the random effects $\boldsymbol{\theta}$ is equal to

$$\Pr(Y_{ij} = c \mid \boldsymbol{\theta}) = P_{ijc} - P_{ij,c-1}\,,$$

where $P_{ijc} = 1/[1+\exp(-\eta_{ijc})]$ under the logistic response function (formulas for other response functions are given in section 6.2.3). Note that because $\gamma_0 = -\infty$ and $\gamma_C = \infty$, $P_{ij0} = 0$ and $P_{ijC} = 1$. Here, η_{ijc} denotes the response model, for example,

$$\eta_{ijc} = \gamma_c - \left[(\boldsymbol{x}^*_{ij})'\boldsymbol{\beta}_c + \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{T}\boldsymbol{\theta}_i\right]\,,$$

or one of the other variants of η_{ijc} presented. In what follows, we'll consider the general model allowing for non-proportional odds, since the more restrictive proportional odds model is just a special case (i.e., when $\boldsymbol{\beta}_c = 0$).

Let $\boldsymbol{Y}_j$ denote the vector of ordinal responses from level-2 unit j (for the n_j level-1 units nested within). The probability of any pattern $\boldsymbol{Y}_j$ conditional on $\boldsymbol{\theta}$ is equal to the product of the probabilities of the level-1 responses,

$$\ell(\boldsymbol{Y}_j \mid \boldsymbol{\theta}) = \prod_{i=1}^{n_j} \prod_{c=1}^{C} (P_{ijc} - P_{ij,c-1})^{\mathrm{y}_{ijc}}, \tag{6.8}$$

where $\mathrm{y}_{ijc} = 1$ if $Y_{ij} = c$ and 0 otherwise (i.e., for each ij-th observation, $\mathrm{y}_{ijc} = 1$ for only one of the C categories). For the ordinal representation of the survival model, where right-censoring is present, the above likelihood is generalized to

$$\ell(\boldsymbol{Y}_j \mid \boldsymbol{\theta}) = \prod_{i=1}^{n_j} \prod_{c=1}^{C} \left[(P_{ijc} - P_{ij,c-1})^{d_{ij}}\,(1 - P_{ijc})^{1-d_{ij}}\right]^{\mathrm{y}_{ijc}}, \tag{6.9}$$

where $d_{ij} = 1$ if Y_{ij} represents an event, or $d_{ij} = 0$ if Y_{ij} represents a censored observation. Notice that (6.9) is equivalent to (6.8) when $d_{ij} = 1$ for all observations. With right-censoring, because there is essentially one additional response category (for those censored at the last category C), it is $\gamma_{C+1} = \infty$ and so $P_{ij,C+1} = 1$. In this case, parameters γ_c and $\boldsymbol{\beta}_c$ with $c = 1, \ldots, C$ are estimable, otherwise c only goes to $C - 1$.

The marginal density of $\underline{\boldsymbol{Y}}_j$ in the population is expressed as the following integral of the likelihood, $\ell(\cdot)$, weighted by the prior density $g(\cdot)$,

$$h(\boldsymbol{Y}_j) = \int_{\boldsymbol{\theta}} \ell(\boldsymbol{Y}_j \mid \boldsymbol{\theta})\, g(\boldsymbol{\theta})\ \mathrm{d}\boldsymbol{\theta}, \tag{6.10}$$

where $g(\boldsymbol{\theta})$ represents the multivariate standard normal density. The marginal log-likelihood from the N level-2 units, $\log L = \sum_j^N \log h(\boldsymbol{Y}_j)$, is then maximized to yield maximum likelihood estimates. For this, denote the conditional likelihood as ℓ_j and the marginal density as h_j. Differentiating first with respect to the parameters that vary with c, let $\boldsymbol{\alpha}_k$ represent a particular threshold γ_k or regression vector $\boldsymbol{\beta}_k^*$, where $k = 1, \ldots, C$ if right-censoring occurs, otherwise $k = 1, \ldots, C - 1$. Then

$$\frac{\partial \log L}{\partial \boldsymbol{\alpha}_k} = \sum_{j=1}^{N} h_j^{-1}\ \frac{\partial h_j}{\partial \boldsymbol{\alpha}_k}\,,$$

with

$$\frac{\partial h_j}{\partial \boldsymbol{\alpha}_k} = \int_{\boldsymbol{\theta}} \sum_{i=1}^{n_j} \sum_{c=1}^{C} \mathrm{y}_{ijc} \left[d_{ij} \frac{(\partial P_{ijc}) a_{ck} - (\partial P_{ij,c-1}) a_{c-1,k}}{P_{ijc} - P_{ij,c-1}} \right.$$
$$\left. - (1 - d_{ij}) \frac{(\partial P_{ijc}) a_{ck}}{1 - P_{ijc}} \right] \times \ell_j\, g(\boldsymbol{\theta})\, \frac{\partial \eta_{ijk}}{\partial \boldsymbol{\alpha}_k}\ \mathrm{d}\boldsymbol{\theta}, \tag{6.11}$$

where $\partial \eta_{ijk} / \partial \boldsymbol{\alpha}_k = 1$ and $-\boldsymbol{x}_{ij}^*$ for the thresholds and regression coefficients, respectively, and $a_{ck} = 1$ if $c = k$ (and $= 0$ if $c \neq k$). Also, ∂P_{ijc} represents the pdf of the response function; various forms of this are given in section 6.2.3.

For the parameters that do not vary with c, let $\boldsymbol{\zeta}$ represent an arbitrary parameter vector; then for $\boldsymbol{\beta}$ and the vector $\mathrm{v}(\boldsymbol{T})$, which contains the unique elements of the Cholesky factor $\boldsymbol{T}$, we get

$$\frac{\partial \log L}{\partial \boldsymbol{\zeta}} = \sum_{j=1}^{N} h_j^{-1} \int_{\boldsymbol{\theta}} \sum_{i=1}^{n_j} \sum_{c=1}^{C} \mathrm{y}_{ijc} \left[d_{ij} \frac{\partial P_{ijc} - \partial P_{ij,c-1}}{P_{ijc} - P_{ij,c-1}} - (1 - d_{ij}) \frac{\partial P_{ijc}}{1 - P_{ijc}} \right]$$
$$\times\, \ell_j\, g(\boldsymbol{\theta})\, \frac{\partial \eta_{ijc}}{\partial \boldsymbol{\zeta}}\ \mathrm{d}\boldsymbol{\theta}, \tag{6.12}$$

where

$$\frac{\partial \eta_{ijc}}{\partial \boldsymbol{\beta}} = -\boldsymbol{x}_{ij}\,, \qquad \frac{\partial \eta_{ijc}}{\partial (\mathrm{v}(\boldsymbol{T}))} = -\mathsf{J}_r(\boldsymbol{\theta} \otimes \boldsymbol{z}_{ij}),$$

and J_r is the elimination matrix of Magnus [69], which eliminates the elements above the main diagonal. If $\boldsymbol{T}$ is an $r \times 1$ vector of independent variance terms (e.g., if $\boldsymbol{z}_{ij}$ is an $r \times 1$ vector of level-1 or level-2 grouping variables, see section 6.7), then $\partial \eta_{ijc} / \partial \boldsymbol{T} = \boldsymbol{z}_{ij} \theta$ in the equation above.

Fisher's method of scoring can be used to provide the solution to these likelihood equations. For this, provisional estimates for the vector of parameters $\boldsymbol{\Theta}$, on iteration ι are improved by

$$\boldsymbol{\Theta}_{\iota+1} = \boldsymbol{\Theta}_\iota - \left\{ E \left[\frac{\partial^2 \log L}{\partial \boldsymbol{\Theta}_\iota \, \partial \boldsymbol{\Theta}'_\iota} \right] \right\}^{-1} \frac{\partial \log L}{\partial \boldsymbol{\Theta}_\iota} \,, \tag{6.13}$$

where, following Bock and Lieberman [17], the information matrix, or minus the expectation of the matrix of second derivatives, is given by

$$- E \left[\frac{\partial^2 \log L}{\partial \boldsymbol{\Theta}_\iota \, \partial \boldsymbol{\Theta}'_\iota} \right] = E \left[\sum_{j=1}^{N} h_j^{-2} \, \frac{\partial h_j}{\partial \boldsymbol{\Theta}_\iota} \left(\frac{\partial h_j}{\partial \boldsymbol{\Theta}_\iota} \right)' \right] .$$

Its estimator is obtained using the estimated parameter values and, at convergence, the large-sample variance covariance matrix of the parameter estimates is gotten as the inverse of the information matrix. The form on the right-hand side of the above equation is sometimes called the "outer product of the gradients." It was proposed in the econometric literature by Berndt et al. [12], and is often referred to as the BHHH method.

6.4 Multilevel Nominal Response Models

Let $\underline{Y}_{ij}$ now denote a nominal variable associated with level-2 unit j and level-1 unit i. Adding random effects to the fixed-effects multinomial logistic regression model (see Agresti [2], Long [65]), we get that the probability that $\underline{Y}_{ij} = c$ (a response occurs in category c) for a given level-2 unit j is given by

$$\underline{p}_{ijc} = \Pr(\underline{Y}_{ij} = c) = \frac{\exp(\underline{\eta}_{ijc})}{1 + \sum_{h=2}^{C} \exp(\underline{\eta}_{ijh})} \qquad \text{for } c = 2, 3, \ldots, C, \tag{6.14a}$$

$$\underline{p}_{ij1} = \Pr(\underline{Y}_{ij} = 1) = \frac{1}{1 + \sum_{h=2}^{C} \exp(\underline{\eta}_{ijh})} \,, \tag{6.14b}$$

where the multinomial logit $\underline{\eta}_{ijc} = \boldsymbol{x}'_{ij} \boldsymbol{\beta}_c + \boldsymbol{z}'_{ij} \boldsymbol{T}_c \, \underline{\boldsymbol{\theta}}_j$. Comparing this to the logit for ordered responses, we see that all of the covariate effects $\boldsymbol{\beta}_c$ vary across categories ($c = 2, 3, \ldots, C$). Similarly for the random-effect variance term $\boldsymbol{T}_c$. As written above, an important distinction between the model for ordinal and nominal responses is that the former uses cumulative comparisons of the categories whereas the latter uses comparisons to a reference category.

This model generalizes Bock's model for educational test data [14] by including covariates $\boldsymbol{x}_{ij}$, and by allowing a general random-effects design vector $\boldsymbol{z}_{ij}$ including the possibility of multiple random effects $\underline{\boldsymbol{\theta}}_j$. As discussed by Bock [14], the model has a plausible interpretation. Namely, each nominal category is assumed to be related to an underlying latent "response tendency" for that category. The category c associated with the response variable $\underline{Y}_{ij}$ is then the category for which the response tendency is maximal. Notice that this assumption of C latent variables differs from the ordinal model where only one underlying latent variable is assumed. Bock [15] refers to the former as the extremal concept and the latter as the aforementioned threshold concept, and notes that both were introduced into psychophysics by Thurstone [111]. The two are equivalent only for the dichotomous case (i.e., when there are only two response categories).

The model as written above allows estimation of any pairwise comparisons among the C response categories. As characterized in Bock [14], it is benefical to write the nominal model to allow for any possible set of $C-1$ contrasts. For this, the category probabilities are written as

$$\underline{p}_{ijc} = \frac{\exp(\underline{\eta}_{ijc})}{\sum_{h=1}^{C} \exp(\underline{\eta}_{ijh})} \qquad \text{for } c = 1, 2, \ldots, C, \tag{6.15}$$

where now

$$\underline{\eta}_{ijc} = \boldsymbol{x}'_{ij} \boldsymbol{\Gamma} \boldsymbol{d}_c + (\boldsymbol{z}'_{ij} \otimes \underline{\boldsymbol{\theta}}'_j) \mathsf{J}'_{r^*} \boldsymbol{\Lambda} \boldsymbol{d}_c \,. \tag{6.16}$$

Here, $\boldsymbol{D}$ is the $(C-1) \times C$ matrix containing the contrast coefficients for the $C-1$ contrasts between the C logits and $\boldsymbol{d}_c$ is the cth column vector of this matrix. The $s \times (C-1)$ parameter matrix $\boldsymbol{\Gamma}$ contains the regression coefficients associated with the s covariates for each of the $C-1$ contrasts. Similarly, $\boldsymbol{\Lambda}$ contains the random-effect variance parameters for each of the $C-1$ contrasts. Specifically,

$$\boldsymbol{\Lambda} = [\ \mathrm{v}(\boldsymbol{T}_1) \quad \mathrm{v}(\boldsymbol{T}_2) \quad \ldots \quad \mathrm{v}(\boldsymbol{T}_{C-1})\ ] \,,$$

where $\mathrm{v}(\boldsymbol{T}_c)$ is the $r^* \times 1$ vector ($r^* = r(r+1)/2$) of elements below and on the diagonal of the Cholesky (lower-triangular) factor $\boldsymbol{T}_c$, and J_{r^*} is the aforementioned elimination matrix of Magnus [69]. This latter matrix is necessary to ensure that the appropriate terms from the $1 \times r^2$ vector resulting from the Kronecker product $(\boldsymbol{z}'_{ij} \otimes \underline{\boldsymbol{\theta}}'_j)$ are multiplied with the $r^* \times 1$ vector resulting from $\boldsymbol{\Lambda d}_c$. For the case of a random-intercepts model, the model simplifies to

$$\underline{\eta}_{ijc} = \boldsymbol{x}'_{ij} \boldsymbol{\Gamma} \boldsymbol{d}_c + \boldsymbol{\Lambda} \boldsymbol{d}_c \, \underline{\theta}_j \,,$$

with $\boldsymbol{\Lambda}$ as the $1 \times (C-1)$ vector $\boldsymbol{\Lambda} = [\ \sigma_1 \quad \sigma_2 \quad \ldots \quad \sigma_{C-1}\ ]$.

Notice that if $\boldsymbol{D}$ equals

$$\boldsymbol{D} = \begin{pmatrix} 0\,1\,0\ldots0 \\ 0\,0\,1\ldots0 \\ .\;.\;.\;.\;.\;.\;.\;. \\ 0\,0\,0\ldots1 \end{pmatrix},$$

the model simplifies to the earlier representation in (6.14a) and (6.14b). The current formulation, however, allows for a great deal of flexibility in the types of comparisons across the C response categories. For example, if the categories are ordered, an alternative to the cumulative logit model of the previous section is to employ Helmert contrasts [15] within the nominal model. For this, with $C = 4$, the contrast matrix would be

$$\boldsymbol{D} = \begin{pmatrix} -1 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & -1 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & -1 & 1 \end{pmatrix}.$$

Helmert contrasts are similar to the category comparisons of continuation-ratio logit models, as described within a mixed model formulation by Ten Have and Uttal [108]. However, the Helmert contrasts above are applied to the category logits, rather then the category probabilities as in continuation-ratio models.

6.4.1 Parameter Estimation

Estimation follows the procedure described for ordinal outcomes. Specifically, letting $\boldsymbol{Y}_j$ denote the vector of nominal responses from level-2 unit j (for the n_j level-1 units nested within), the probability of any $\boldsymbol{Y}_j$ conditional on the random effects $\boldsymbol{\theta}$ is equal to the product of the probabilities of the level-1 responses

$$\ell(\boldsymbol{Y}_j \mid \boldsymbol{\theta}) = \prod_{i=1}^{n_j} \prod_{c=1}^{C} (p_{ijc})^{\mathrm{y}_{ijc}}, \tag{6.17}$$

where $\mathrm{y}_{ijc} = 1$ if $Y_{ij} = c$, and 0 otherwise. The marginal density of the response vector $\boldsymbol{Y}_j$ is again given by (6.10). The marginal log-likelihood from the N level-2 units, $\log L = \sum_j^N \log h(\boldsymbol{Y}_j)$, is maximized to obtain maximum likelihood estimates of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Lambda}$. Specifically, using $\boldsymbol{\Delta}$ to represent either parameter matrix,

$$\frac{\partial \log L}{\partial \boldsymbol{\Delta}'} = \sum_{j=1}^{N} h^{-1}(\boldsymbol{Y}_j) \int_{\boldsymbol{\theta}} \left[\sum_{i=1}^{n_j} \boldsymbol{D} \left(\mathbf{y}_{ij} - \boldsymbol{P}_{ij} \right) \otimes \partial \boldsymbol{\Delta} \right] \times \ell(\boldsymbol{Y}_j \mid \boldsymbol{\theta})\; g(\boldsymbol{\theta})\; \mathrm{d}\boldsymbol{\theta}, \tag{6.18}$$

where

$$\partial \boldsymbol{\Gamma} = \boldsymbol{x}'_{ij} \,, \qquad \partial \boldsymbol{\Lambda} = [\mathsf{J}_{r^*}(\boldsymbol{\theta} \otimes \boldsymbol{z}_{ij})]' \,,$$

$\mathbf{y}_{ij}$ is the $C \times 1$ indicator vector, and $\boldsymbol{P}_{ij}$ is the $C \times 1$ vector obtained by applying (6.15) for each category. As in the ordinal case, Fisher's method of scoring can be used to provide the solution to these likelihood equations.

6.5 Computational Issues

In order to solve the above likelihood solutions for both the ordinal and nominal models, integration over the random-effects distribution must be performed. Additionally, the above likelihood solutions are only in terms of the regression parameters and variance-covariance parameters of the random-effects distribution. Often, estimation of the random effects is also of interest. These issues are described in great detail in Skrondal and Rabe-Hesketh [98]; here, we discuss some of the relevant points.

6.5.1 Integration over $\boldsymbol{\theta}$

Various approximations for evaluating the integral over the random-effects distribution have been proposed in the literature; several of these are compared in chapter 9. Perhaps the most frequently used methods are based on first- or second-order Taylor expansions. Marginal quasi-likelihood (MQL) involves expansion around the fixed part of the model, whereas penalized or predictive quasi-likelihood (PQL) additionally includes the random part in its expansion [39]. Both of these are available in the MLwiN software program [84]. Unfortunately, several authors [19, 87, 90] have reported downwardly biased estimates using these procedures in certain situations, especially for the first-order expansions.

Raudenbush et al. [87] proposed an approach that uses a combination of a fully multivariate Taylor expansion and a Laplace approximation. Based on the results in Raudenbush et al. [87], this method yields accurate results and is computationally fast. Also, as opposed to the MQL and PQL approximations, the deviance obtained from this approximation can be used for likelihood-ratio tests. This approach has been incorporated into the HLM software program [86].

Numerical integration can also be used to perform the integration over the random-effects distribution. Specifically, if the assumed distribution is normal, Gauss-Hermite quadrature can approximate the above integral to any practical degree of accuracy [104]. Additionally, like the Laplace approximation, the numerical quadrature approach yields a deviance that can be readily used for likelihood-ratio tests. The integration is approximated by a summation on a specified number of quadrature points Q for each dimension of the integration.

The solution via quadrature can involve summation over a large number of points, especially as the number of random effects is increased. For example, if there is only one random effect, the quadrature solution requires only one additional summation over Q points relative to the fixed effects solution. For models with $r > 1$ random effects, however, the quadrature is performed over Q^r points, and so becomes computationally burdensome for $r > 5$ or so. Also, Lesaffre and Spiessens [61] present an example where the method only gives valid results for a high number of quadrature points. These authors advise practitioners to routinely examine results for the dependence on Q. To address these issues, several authors have described a method of adaptive quadrature that uses relatively few points per dimension (e.g., 3 or so), which are adapted to the location and dispersion of the distribution to be integrated [18, 64, 78, 80]. Simulations show that adaptive quadrature performs well in a wide variety of situations and typically outperforms ordinary quadrature [82]. Several software packages have implemented ordinary or adaptive Gauss-Hermite quadrature, including EGRET [22], GLLAMM [81], LIMDEP [40], MIXOR [49], MIXNO [46], Stata [101], and SAS PROC NLMIXED [93].

Another approach that is commonly used in econometrics and transportation research uses simulation methods to integrate over the random-effects distribution (see the introductory overview by Stern [102] and the excellent book by Train [112]). When used in conjunction with maximum likelihood estimation, it is called "maximum simulated likelihood" or "simulated maximum likelihood." The idea behind this approach is to draw a number of values from the random-effects distribution, calculate the likelihood for each of these draws, and average over the draws to obtain a solution. Thus, this method maximizes a simulated sample likelihood instead of an exact likelihood, but can be considerably faster than quadrature methods, especially as the number of random effects increases [41]. It is a very flexible and intuitive approach with many potential applications (see Drukker [27]). In particular, Bhat [13] and Glasgow [36] describe this estimation approach for multilevel models of nominal outcomes. In terms of software, LIMDEP [40] has included this estimation approach for several types of outcome variables, including nominal and ordinal, and Haan and Uhlendorff [41] describe a Stata routine for nominal data.

Bayesian approaches, such as the use of Gibbs sampling [33] and related methods [105], can also be used to integrate over the random effects distribution. This approach is described in detail in chapter 2. For nominal responses, Daniels and Gatsonis [25] use this approach in their multilevel polychotomous regression model. Similarly, Ishwaran [57] utilize Bayesian methods in modeling multilevel ordinal data. The freeware BUGS software program [100] can be used to facilitate estimation via Gibbs sampling. In this regard, Marshall and Spiegelhalter [70] provides an example of multilevel modeling using BUGS, including some syntax and discussion of the program.

6.5.2 Estimation of Random Effects and Probabilities

In many cases, it is useful to obtain estimates of the random effects and also to obtain fitted marginal probabilities. The random effects $\boldsymbol{\theta}_j$ can be estimated using empirical Bayes methods [16]. For the univariate case, this estimator $\hat{\theta}_j$ is given by the mean of the posterior distribution,

$$\hat{\theta}_j = E(\underline{\theta}_j \mid \mathbf{Y}_j) = \frac{1}{h(\mathbf{Y}_j)} \int_\theta \theta_j \; \ell(\cdot) \; g(\theta) \; \mathrm{d}\theta, \tag{6.19}$$

where $\ell(\cdot)$ is the conditional likelihood for the particular model (i.e., ordinal or nominal). The variance of the posterior distribution is obtained as

$$\mathrm{Var}(\hat{\underline{\theta}}_j \mid \mathbf{Y}_j) = \frac{1}{h(\mathbf{Y}_j)} \int_\theta (\theta_j - \hat{\theta}_j)^2 \; \ell(\cdot) \; g(\theta) \; \mathrm{d}\theta.$$

These quantities may be used, for example, to evaluate the response probabilities for particular level-2 units (e.g., person-specific trend estimates).

To obtain estimated marginal probabilities (e.g., the estimated response probabilities of the control group across time), an additional step is required for models with non-linear response functions (e.g., the models considered in this paper). First, so-called “subject-specific” probabilities [75, 118] are estimated for specific values of covariates and random effects, say θ^*. These subject-specific estimates indicate, for example, the response probability for a subject with random effect level θ^* in the control group at a particular timepoint. Denoting these subject-specific probabilities as $\hat{P}_{ss}$, marginal probabilities $\hat{P}_m$ can then be obtained by numerical quadrature, namely $\hat{P}_m = \int_\theta \hat{P}_{ss} \; g(\theta) \; \mathrm{d}\theta$, or by marginalizing the scale of the regression coefficients [51, p. 179]. Continuing with our example, the marginalized estimate would indicate the estimated response probability for the entire control group at a particular timepoint. Both subject-specific and marginal estimates have their uses, since they are estimating different quantities, and several authors have characterized the differences between the two [45, 62, 75].

6.6 Intraclass Correlation

For a random-intercepts model (i.e., $\boldsymbol{z}_j = \mathbf{1}_{n_j}$) it is often of interest to express the level-2 variance in terms of an intraclass correlation. For this, one can make reference to the threshold concept and the underlying latent response tendency that determines the observed response. For the ordinal logistic model assuming normally distributed random-effects, the estimated intraclass correlation equals $\hat{\sigma}^2/(\hat{\sigma}^2 + \pi^2/3)$, where the latter term in the denominator represents the variance of the underlying latent response tendency. As mentioned earlier, for the logistic model, this variable is assumed to be distributed as a standard logistic distribution with variance equal to $\pi^2/3$. For a probit model this term is replaced by 1, the variance of the standard normal distribution.

For the nominal model, one can make reference to multiple underlying latent response tendencies, denoted as $\underline{y}_{ijc}$, and the associated regression model including level-1 residuals $\underline{\epsilon}_{ijc}$

$$\underline{y}_{ijc} = \boldsymbol{x}'_{ij}\boldsymbol{\beta}_c + \boldsymbol{z}'_{ij}\boldsymbol{T}_c\,\underline{\boldsymbol{\theta}}_j + \underline{\epsilon}_{ijc} \qquad c = 1, 2, \ldots, C.$$

As mentioned earlier, for a particular ij-th unit, the category c associated with the nominal response variable $\underline{Y}_{ij}$ is the one for which the latent $\underline{y}_{ijc}$ is maximal. Since, in the common reference cell formulation, $c = 1$ is the reference category, $\boldsymbol{T}_1 = \boldsymbol{\beta}_1 = 0$, and so the model can be rewritten as

$$\underline{y}_{ijc} = \boldsymbol{x}'_{ij}\boldsymbol{\beta}_c + \boldsymbol{z}'_{ij}\boldsymbol{T}_c\,\underline{\boldsymbol{\theta}}_j + (\underline{\epsilon}_{ijc} - \underline{\epsilon}_{ij1}) \qquad c = 2, \ldots, C,$$

for the latent response tendency of category c relative to the reference category. It can be shown that the level-1 residuals $\underline{\epsilon}_{ijc}$ for each category are distributed according to a type I extreme-value distribution [see 68, p. 60]. It can further be shown that the standard logistic distribution is obtained as the difference of two independent type I extreme-value variates [see 72, pp. 20 and 142]. As a result, the level-1 variance is given by $\pi^2/3$, which is the variance for a standard logistic distribution. The estimated intraclass correlations are thus calculated as $r_c = \hat{\sigma}_c^2/(\hat{\sigma}_c^2 + \pi^2/3)$, where $\hat{\sigma}_c^2$ is the estimated level-2 variance assuming normally-distributed random intercepts. Notice that $C - 1$ intraclass correlations are estimated, one for each category c versus the reference category. As such, the cluster influence on the level-1 responses is allowed to vary across the nominal response categories.

6.7 Heterogeneous Variance Terms

Allowing for separate random-effect variance terms for groups of either i or j units is sometimes important. For example, in a twin study it is often necessary to allow the intra-twin correlation to differ between monozygotic and dizygotic twins. In this situation, subjects ($i = 1, 2$) are nested within twin pairs ($j = 1, \ldots, N$). To allow the level-2 variance to vary for these two twin-pair types, the random-effects design vector $\boldsymbol{z}_{ij}$ is specified as a 2×1 vector of dummy codes indicating monozygotic and dizygotic twin pair status, respectively. $\boldsymbol{T}$ (or $\boldsymbol{T}_c$ in the nominal model) is then a 2×1 vector of independent random-effect standard deviations for monozygotics and dizygotics, and the cluster effect $\underline{\theta}_j$ is a scalar that is pre-multiplied by the vector $\boldsymbol{T}$. For example, for a random-intercepts proportional odds model, we would have

$$\log\left[\frac{\underline{P}_{ijc}}{1 - \underline{P}_{ijc}}\right] = \gamma_c - \left\{\boldsymbol{x}'_{ij}\boldsymbol{\beta} + [MZ_j \quad DZ_j]\begin{bmatrix}\sigma_{\delta\,(MZ)} \\ \sigma_{\delta\,(DZ)}\end{bmatrix}\underline{\theta}_j\right\},$$

where MZ_j and DZ_j are dummy codes indicating twin pair status (i.e., if $MZ_j = 1$ then $DZ_j = 0$, and vice versa).

Notice, that if the probit formulation is used and the model has no covariates (i.e., only an intercept, $x_{ij} = 1$), the resulting intraclass correlations

$$ICC_{MZ} = \frac{\sigma^2_{\delta\,(MZ)}}{\sigma^2_{\delta\,(MZ)} + 1} \qquad \text{and} \qquad ICC_{DZ} = \frac{\sigma^2_{\delta\,(DZ)}}{\sigma^2_{\delta\,(DZ)} + 1}$$

are polychoric correlations (for ordinal responses) or tetrachoric correlations (for binary responses) for the within twin-pair data. Adding covariates then yields adjusted tetrachoric and polychoric correlations. Because estimation of polychoric and tetrachoric correlations is often important in twin and genetic studies, these models are typically formulated in terms of the probit link. Comparing models that allow homogeneous versus heterogeneous subgroup random-effects variance, thus allows testing of whether the tetrachoric (or polychoric) correlations are equal across the subgroups.

The use of heterogeneous variance terms can also be found in some item response theory (IRT) models in the educational testing literature [14, 16, 92]. Here, item responses ($i = 1, 2, \ldots, m$) are nested within subjects ($j = 1, 2, \ldots, N$) and a separate random-effect standard deviation (i.e., an element of the $m \times 1$ vector $\boldsymbol{T}$) is estimated for each test item (i.e., each i unit). In the multilevel model this is accomplished by specifying $\boldsymbol{z}_{ij}$ as an $m \times 1$ vector of dummy codes indicating the repeated items. To see this, consider the popular two-parameter logistic model for dichotomous responses [66] that specifies the probability of a correct response to item i ($\underline{Y}_{ij} = 1$) as a function of the ability of subject j ($\underline{\theta}_j$),

$$\Pr(\underline{Y}_{ij} = 1) = \frac{1}{1 + \exp[-a_i(\underline{\theta}_j - b_i)]} ,$$

where a_i is the slope parameter for item i (i.e., item discrimination), and b_i is the threshold or difficulty parameter for item i (i.e., item difficulty). The distribution of ability in the population of subjects is assumed to be normal with mean 0 and variance 1 (i.e., the usual assumption for the random effects $\underline{\theta}_j$ in the multilevel model). As noted by Bock and Aitkin [16], it is convenient to let $c_i = -a_i b_i$ and write

$$\Pr(\underline{Y}_{ij} = 1) = \frac{1}{1 + \exp[-(c_i + a_i\underline{\theta}_j)]} ,$$

which can be recast in terms of the logit of the response as

$$\text{logit}_{ij} = \log\left[\frac{\underline{p}_{ij}}{1 - \underline{p}_{ij}}\right] = c_i + a_i\underline{\theta}_j \,.$$

As an example, suppose that there are four items. This model can be represented in matrix form as

$$\begin{pmatrix} \text{logit}_{1j} \\ \text{logit}_{2j} \\ \text{logit}_{3j} \\ \text{logit}_{4j} \end{pmatrix} = \underbrace{\begin{pmatrix} 1\,0\,0\,0 \\ 0\,1\,0\,0 \\ 0\,0\,1\,0 \\ 0\,0\,0\,1 \end{pmatrix}}_{\boldsymbol{X}_j} \underbrace{\begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix}}_{\boldsymbol{c}} + \underbrace{\begin{pmatrix} 1\,0\,0\,0 \\ 0\,1\,0\,0 \\ 0\,0\,1\,0 \\ 0\,0\,0\,1 \end{pmatrix}}_{\boldsymbol{Z}_j} \underbrace{\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix}}_{\boldsymbol{a}} (\underline{\theta}_j),$$

showing that this IRT model is a multilevel model that allows the random effect variance terms to vary across items (level-1). The usual IRT notation is a bit different than the multilevel notation, but $\boldsymbol{c}$ simply represents the fixed-effects (i.e., $\boldsymbol{\beta}$) and $\boldsymbol{a}$ is the the random-effects standard deviation vector $\boldsymbol{T}' = [\ \sigma_{\delta 1} \quad \sigma_{\delta 2} \quad \sigma_{\delta 3} \quad \sigma_{\delta 4}\ ]$.

The elements of the $\boldsymbol{T}$ vector can also be viewed as the (unscaled) factor loadings of the items on the (unidimensional) underlying ability variable ($\underline{\theta}$). A simpler IRT model that constrains these factor loadings to be equal is the one-parameter logistic model, the so-called Rasch model [116]. This constraint is achieved by setting $\boldsymbol{Z}_j = \mathbf{1}_{n_j}$ and $\boldsymbol{a} = a$ in the above model. Thus, the Rasch model is simply a random-intercepts logistic regression model with item indicators for $\boldsymbol{X}$.

Unlike traditional IRT models, the multilevel formulation of the model easily allows multiple covariates at either level (i.e., items or subjects). This and other advantages of casting IRT models as multilevel models are described in detail by Adams et al. [1] and Rijmen et al. [89]. In particular, this allows a model for examining whether item parameters vary by subject characteristics, and also for estimating ability in the presence of such item by subject interactions. Interactions between item parameters and subject characteristics, often termed item bias [20], is an area of active psychometric research. Also, although the above illustration is in terms of a dichotomous response model, the analogous multilevel ordinal and nominal models apply. For ordinal items responses, application of the cumulative logit multilevel models yields what Thissen and Steinberg [110] have termed "difference models," namely, the treatment of ordinal responses as developed by Samejima [92] within the IRT-context. Similarly, in terms of nominal responses, the multilevel model yields the nominal IRT model developed by Bock [14].

6.8 Health Services Research Example

The McKinney Homeless Research Project (MHRP) study [55, 56] in San Diego, CA was designed to evaluate the effectiveness of using section 8 certificates as a means of providing independent housing to the severely mentally ill homeless. Section 8 housing certificates were provided from the Department of Housing and Urban Development (HUD) to local housing authorities in San Diego. These housing certificates, which require clients to pay 30% of their income toward rent, are designed to make it possible for low income individuals to choose and obtain independent housing in the community. Three hundred sixty-one clients took part in this longitudinal study employing a randomized factorial design. Clients were randomly assigned to one of two types of supportive case management (comprehensive vs. traditional) and to one of two levels of access to independent housing (using section 8 certificates). Eligibility for the project was restricted to individuals diagnosed with a severe and persistent mental illness who were either homeless or at high risk of becoming homeless at the start of the study. Individuals' housing status was classified at baseline and at 6, 12, and 24 month follow-ups.

In this illustration, focus will be on examining the effect of access to section 8 certificates on repeated housing outcomes across time. Specifically, at each timepoint each subjects' housing status was classified as either streets/shelters, community housing, or independent housing. This outcome can be thought of as ordinal with increasing categories indicating improved housing outcomes. The observed sample sizes and response proportions for these three outcome categories by group are presented in Table 6.1.

Table 6.1. Housing status across time by group: response proportions and sample sizes.

		timepoint			
group	status	baseline	6-months	12-months	24-months
control	street	.555	.186	.089	.124
	community	.339	.578	.582	.455
	independent	.106	.236	.329	.421
	n	180	161	146	145
section 8	street	.442	.093	.121	.120
	community	.414	.280	.146	.228
	independent	.144	.627	.732	.652
	n	181	161	157	158

These observed proportions indicate a general decrease in street living and an increase in independent living across time for both groups. The increase in independent housing, however, appears to occur sooner for the section 8 group relative to the control group. Regarding community living, across time this increases for the control group and decreases for the section 8 group.

There is some attrition across time; attrition rates of 19.4% and 12.7% are observed at the final timepoint for the control and section 8 groups, respectively. Since estimation of model parameters is based on a full-likelihood approach, the missing data are assumed to be "ignorable" conditional on both the model covariates and the observed responses [60]. In longitudinal studies, ignorable nonresponse falls under the "missing at random" (MAR) assumption introduced by Rubin [91], in which the missingness depends only on observed data. In what follows, since the focus is on describing application of the various multilevel regression models, we will make the MAR assumption. A further approach, however, that does not rely on the MAR assumption (e.g., a multilevel pattern-mixture model as described in Hedeker and Gibbons [50]) could be used. Missing data issues are described more fully in chapter 10.

6.8.1 Ordinal Response Models

To prepare for the ordinal analyses, the observed cumulative logits across time for the two groups are plotted in Figures 6.3 and 6.4 The first cumulative logit compares independent and community housing versus street living (i.e., categories 2 & 3 combined versus 1), while the second cumulative logit compares independent housing versus community housing and street living (i.e., category 3 versus 2 and 1 combined). For the proportional odds model to hold, these two plots should look the same, with the only difference being the scale difference on the y-axis. As can be seen, these plots do not look that similar. For example, the post-baseline group differences do not appear to be the same for the two cumulative logits. In particular, it appears that the section 8 group does better more consistently in terms of the second cumulative logit (i.e., independent versus community and street housing). This would imply that the proportional odds model is not reasonable for these data.

To assess this more rigorously, two ordinal logistic multilevel models were fit to these data, the first assuming a proportional odds model and the second relaxing this assumption. For both analyses, the repeated housing status classifications were modeled in terms of time effects (6, 12, and 24 month follow-ups compared to baseline), a group effect (section 8 versus control), and group by time interaction terms. The first analysis assumes these effects are the same across the two cumulative logits of the model, whereas the second analysis estimates effects for each explanatory variable on each of the two cumulative logits. In terms of the multilevel part of the model, only a random subject effect was included in both analyses. Results from these analyses are given in Table 6.2.

The proportional odds model indicates significant time effects for all timepoints relative to baseline, but only significant group by time interactions for the 6 and 12 month follow-ups. Marginally significant effects are obtained for the section 8 effect and the section 8 by t3 (24-months) interaction. Thus,

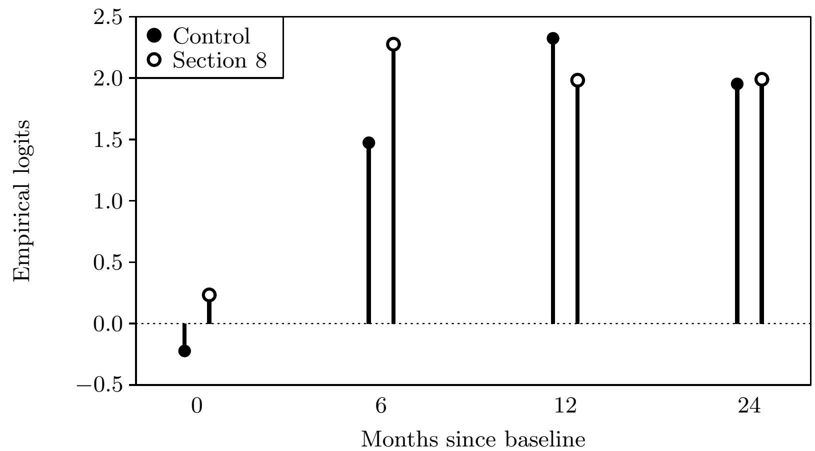

Fig. 6.3. First cumulative logit values across time by group.

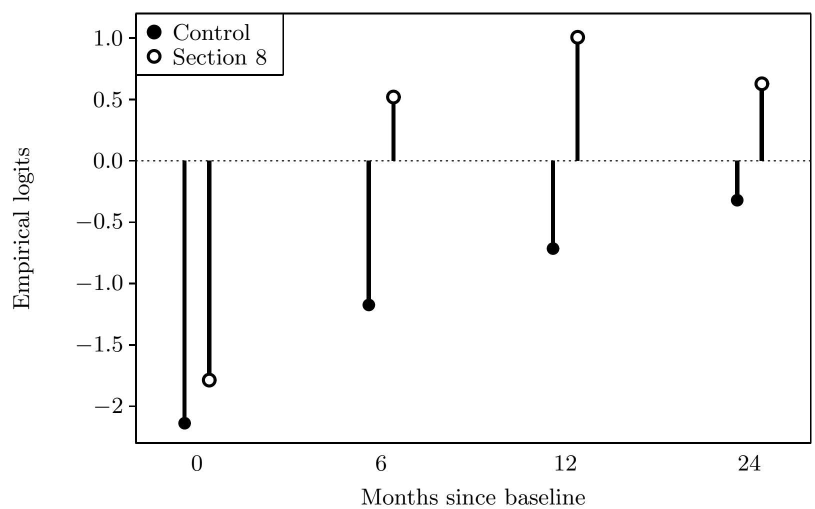

Fig. 6.4. Second cumulative logit values across time by group.

the analysis indicates that the control group moves away from street living to independent living across time, and that this improvement is more pronounced for section 8 subjects at the 6 and 12 month follow-up. Because the section 8 by t3 interaction is only marginally significant, the groups do not differ significantly in housing status at the 24-month follow-up as compared to baseline.

However, comparing log-likelihood values clearly rejects the proportional odds assumption (likelihood ratio $\chi^2_7 = 52.14$) indicating that the effects of the explanatory variables cannot be assumed identical across the two cumu-

Table 6.2. Housing status across time: Ordinal logistic model estimates and standard errors (se).

	Proportional Odds		Non-Proportional Odds			
			Non-street[1]		Independent[2]	
term	estimate	se	estimate	se	estimate	se
intercept	−.220	.203	−.322	.218		
threshold	**2.744**	.110			**2.377**	.279
t1 (6 month vs base)	**1.736**	.233	**2.297**	.298	**1.079**	.358
t2 (12 month vs base)	**2.315**	.268	**3.345**	.450	**1.645**	.336
t3 (24 month vs base)	**2.499**	.247	**2.821**	.369	**2.145**	.339
section 8 (yes=1, no=0)	*.497*	.280	*.592*	.305	.323	.401
section 8 by t1	**1.408**	.334	.566	.478	**2.023**	.478
section 8 by t2	**1.173**	.360	−.958	.582	**2.016**	.466
section 8 by t3	*.638*	.331	−.366	.506	**1.073**	.472
subject sd	**1.459**	.106	**1.457**	.112		
$-2 \log L$	2274.39		2222.25			

bold indicates $p < .05$, *italic* indicates $.05 < p < .10$
[1] logit comparing independent and community housing vs. street
[2] logit comparing independent housing vs. community housing and street

lative logits. Interestingly, none of the section 8 by time interaction terms are significant in terms of the non-street logit (i.e., comparing categories 2 and 3 versus 1), while all of them are significant in terms of the independent logit (i.e., comparing category 3 versus 1 and 2 combined). Thus, as compared to baseline, section 8 subjects are more likely to be in independent housing at all follow-up timepoints, relative to the control group.

In terms of the random subject effect, it is clear that the data are correlated within subjects. Expressed as an intra-class correlation, the attributable variance at the subject-level equals .39 for both models. Also, the Wald test is highly significant in terms of rejecting the null hypothesis that the (subject) population standard deviation equals zero. Strictly speaking, as noted by Raudenbush and Bryk [85] and others, this test is not to be relied upon, especially as the population variance is close to zero. In the present case, the actual significance test is not critical because it is more or less assumed that the population distribution of the subject effects will not have zero variance.

6.8.2 Nominal Response Models

For the initial set of analyses with nominal models, reference category contrasts were used and street/shelter was chosen as the reference category. Thus, the first comparison compares community to street responses, and the second compares independent to street responses. A second analysis using Helmert contrasts will be described later.

Corresponding observed logits for the reference-cell comparisons by group and time are given in Figures 6.5 and 6.6. Comparing these plots, different patterns for the post-baseline group differences are suggested. It seems that the non-section 8 group does better in terms of the community versus street comparison, whereas the section 8 group is improved for the independent versus street comparison. Further, the group differences appear to vary across time. The subsequent analyses will examine these visual impressions of the data.

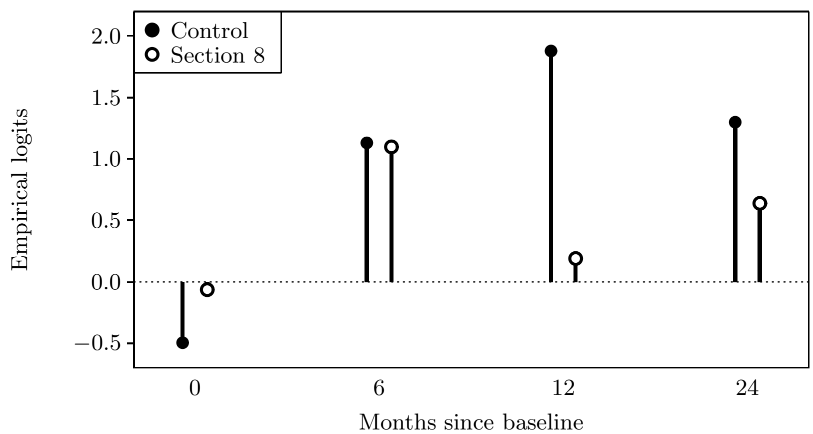

Fig. 6.5. First reference-cell logit values across time by group.

To examine the sensivity of the results to the normality assumption for the random effects, two multilevel nominal logistic regression models were fit to these data assuming the random effects were normally and uniformly distributed, respectively. Tables 6.3 and 6.4 list results for the two response category comparisons of community versus street and independent versus street, respectively. The time and group effects are the same as in the previous ordinal analyses.

The results are very similar for the two multilevel models. Thus, the random-effects distributional form does not seem to play an important role for these data. Subjects in the control group increase both independent and community housing relative to street housing at all three follow-ups, as com-

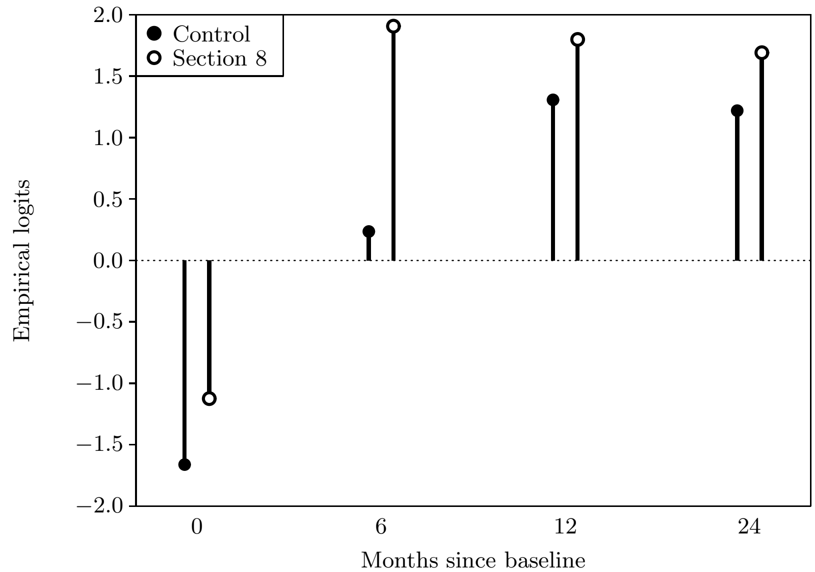

Fig. 6.6. Second reference-cell logit values across time by group.

Table 6.3. Housing status across time: Nominal model estimates and standard errors (se).

Community versus Street	Normal prior		Uniform prior	
term	estimate	se	estimate	se
intercept	**−.452**	.192	**−.473**	.184
t1 (6 month vs base)	**1.942**	.312	**1.850**	.309
t2 (12 month vs base)	**2.820**	.466	**2.686**	.457
t3 (24 month vs base)	**2.259**	.378	**2.143**	.375
section 8 (yes=1, no=0)	*.521*	.268	*.471*	.258
section 8 by t1	−.135	.490	−.220	.484
section 8 by t2	**−1.917**	.611	**−1.938**	.600
section 8 by t3	*−.952*	.535	*−.987*	.527
subject sd	**.871**	.138	**.153**	.031
$-2\log L$	2218.73		2224.74	

bold indicates $p < .05$, *italic* indicates $.05 < p < .10$

Table 6.4. Housing status across time: Nominal model estimates and standard errors (se).

Independent versus Street				
	Normal prior		Uniform prior	
term	estimate	se	estimate	se
intercept	**−2.675**	.367	**−2.727**	.351
t1 (6 month vs base)	**2.682**	.425	**2.540**	.422
t2 (12 month vs base)	**4.088**	.559	**3.916**	.551
t3 (24 month vs base)	**4.099**	.469	**3.973**	.462
section 8 (yes=1, no=0)	.781	.491	.675	.460
section 8 by t1	**2.003**	.614	**2.016**	.605
section 8 by t2	.548	.694	.645	.676
section 8 by t3	.304	.615	.334	.600
subject sd	**2.334**	.196	**.490**	.040
$-2 \log L$	2218.73		2224.74	

bold indicates $p < .05$, *italic* indicates $.05 < p < .10$

pared to baseline. Compared to controls, the increase in community versus street housing is less pronounced for section 8 subjects at 12 months, but not statistically different at 6 months and only marginally different at 24 months. Conversely, as compared to controls, the increase in independent versus street housing is more pronounced for section 8 subjects at 6 months, but not statistically different at 12 or 24 months. Thus, both groups reduce the degree of street housing, but do so in somewhat different ways. The control group subjects are shifted more towards community housing, whereas section 8 subjects are more quickly shifted towards independent housing.

As in the ordinal case, the Wald tests are all significant for the inclusion of the random effects variance terms. A likelihood-ratio test also clearly supports inclusion of the random subject effect (likelihood ratio $\chi^2_2 = 134.3$ and 128.3 for the normal and uniform distribution, respectively, as compared to the fixed-effects model, not shown). Expressed as intraclass correlations, $r_1 = .19$ and $r_2 = .62$ for community versus street and independent versus street, respectively. Thus, the subject influence is much more pronounced in terms of distinguishing independent versus street living, relative to community versus street living. This is borne out by contrasting models with separate versus a common random-effect variance across the two category contrasts (not shown) which yields a highly significant likelihood ratio $\chi^2_1 = 49.2$ favoring the model with separate variance terms.

An analysis was also done to examine if the random-effect variance terms varied significantly by treatment group. The deviance ($-2 \log L$) for this model, assuming normally distributed random effects, equaled 2218.43, which was nearly identical to the value of 2218.73 (from Tables 6.3 and 6.4) for the model assuming homogeneous variances across groups. The control group and section 8 group estimates of the subject standard deviations were respectively .771 (se = .182) and .966 (se = .214) for the community versus street comparison, and 2.228 (se = .299) and 2.432 (se = .266) for the independent versus street comparison. Thus, the homogeneity of variance assumption across treatment groups is clearly not rejected.

Finally, Table 6.5 lists the results obtained for an analysis assuming normally-distributed random effects and using Helmert contrasts for the three response categories. From this analysis, it is interesting that none of the section 8 by time interaction terms are observed to be statistically significant for the first Helmert contrast (i.e., comparing street to non-street housing). Thus, group assignment is not significantly related to housing when considering simply street versus non-street housing outcomes. However, the second Helmert contrast that contrasts the two types of non-street housing (i.e., independent versus community) does reveal the benefical effect of the section 8 certificate in terms of the positive group by time interaction terms. Again, the section 8 group is more associated with independent housing, relative to community housing, than the non-section 8 group. In many ways, the Helmert contrasts, with their intuitive interpretations, represent the best choice for the analysis of these data.

Table 6.5. Housing status across time: Nominal model estimates and standard errors (se) using Helmert contrasts.

	Independent & Community vs Street		Independent vs Community	
term	estimate	se	estimate	se
intercept	**−1.042**	.163	**−1.112**	.163
t1 (6 month vs base)	**1.541**	.215	**.371**	.187
t2 (12 month vs base)	**2.303**	.323	**.634**	.176
t3 (24 month vs base)	**2.119**	.258	**.920**	.179
section 8 (yes=1, no=0)	*.434*	.222	.130	.213
section 8 by t1	*.623*	.330	**1.069**	.253
section 8 by t2	−.457	.401	**1.233**	.256
section 8 by t3	−.216	.345	*.628*	.255
subject sd	**1.068**	.099	**.732**	.083
$-2 \log L = 2218.73$				

bold indicates $p < .05$, *italic* indicates $.05 < p < .10$

6.9 Discussion

Multilevel ordinal and multinomial logistic regression models are described for the analysis of categorical data. These models are useful for analysis of outcomes with more than two response categories. By and large, the models are seen as extensions of the multilevel logistic regression model. However, they generalize the model in different ways. The ordinal model uses cumulative dichotomizations of the categorical outcome. Alternatively, the nominal model typically uses dichotomizations that are based on selecting one category as the reference that the others are each compared to. This chapter has also described how other comparisons can be embedded within the nominal model.

For ordinal data, both proportional odds and non-proportional odds models are considered. Since, as noted by Peterson and Harrell [77], examples of non-proportional odds are not difficult to find, the latter model is especially attractive for analyzing ordinal outcomes. In the example presented, the non-proportional odds model provided more specific information about the effect of section 8 certificates. Namely, as compared to baseline, these certificates were effective in increasing independent housing (versus community housing and street living combined) at all follow-up timepoints. Interestingly, the same could not be said when comparing independent and community housing combined versus street living. Thus, the use of the non-proportional odds model was helpful in elucidating a more focused analysis of the effect of the section 8 program.

For the nominal model, both reference cell and Helmert contrasts were applied in the analysis of these data. The former indicated an increase for community relative to street housing for the non-section 8 group, and an increase for independent relative to street housing for the section 8 group. Alternatively, the Helmert contrasts indicated that the groups did not differ in terms of non-street versus street housing, but did differ in terms of the type of non-street housing (i.e., the section 8 group was more associated with independent housing). In either case, the nominal model makes an assumption that has been referred to as "independence of irrelevant alternatives" [10, 67, 68]. This is because the effect of an explanatory variable comparing two categories is the same regardless of the total number of categories considered. This assumption is generally reasonable when the categories are distinct and dissimilar, and unreasonable as the nominal categories are seen as substitutes for one another [8, 73]. Furthermore, McFadden [74] notes that the multinomial logistic regression model is relatively robust in many cases in which this assumption is implausible. In the present example, the outcome categories are fairly distinct and so the assumption would seem to be reasonable for these data. The possibility of relaxing this assumption, though, for a more general multilevel nominal regression model is discussed in detail in Train [112].

The example presented illustrated the usefulness of the multilevel approach for longitudinal categorical data. In particular, it showed the many possible models and category comparisons that are possible if the response variable has more than two categories. In terms of the multilevel part of the model, only random-intercepts models were considered in the data analysis. However, in describing model development, multiple random effects were allowed. An analysis of these data incorporating random subject intercepts and linear trends is discussed in Hedeker [46]. Additionally, the data had a relatively simple multilevel structure, in that there were only two levels, namely, repeated observations nested within subjects. Extensions of both the ordinal and nominal models for three and higher level is possible in the MLwiN [84] and HLM [86] software programs.

Acknowledgement. Thanks are due to Drs. Richard Hough and Michael Hurlburt for use of the longitudinal data, to Jan de Leeuw for organizing this collection, and especially to Erik Meijer for many helpful comments on previous drafts of this chapter. This work was supported by National Institutes of Mental Health Grant MH56146.

References

1. R. J. Adams, M. Wilson, and M. Wu. Multilevel item response models: An approach to errors in variable regression. *Journal of Educational and Behavioral Statistics*, 22:47–76, 1997.
2. A. Agresti. *Categorical Data Analysis*. Wiley, New York, 1990.
3. A. Agresti, J. G. Booth, J. P. Hobart, and B. Caffo. Random-effects modeling of categorical response data. *Sociological Methodology*, 30:27–80, 2000.
4. A. Agresti and J. B. Lang. A proportional odds model with subject-specific effects for repeated ordered categorical responses. *Biometrika*, 80:527–534, 1993.
5. A. Agresti and R. Natarajan. Modeling clustered ordered categorical data: A survey. *International Statistical Review*, 69:345–371, 2001.
6. P. D. Allison. Discrete-time methods for the analysis of event histories. *Sociological Methodology*, 13:61–98, 1982.
7. P. D. Allison. *Survival Analysis using the SAS System: A Practical Guide*. SAS Institute, Cary, NC, 1995.
8. T. Amemiya. Qualitative response models: A survey. *Journal of Economic Literature*, 19:483–536, 1981.
9. D. A. Anderson and M. Aitkin. Variance component models with binary response: Interviewer variability. *Journal of the Royal Statistical Society, Series B*, 47:203–210, 1985.
10. K. J. Arrow. *Social Choice and Individual Values*. Wiley, New York, 1951.
11. D. J. Bartholomew and M. Knott. *Latent Variable Models and Factor Analysis*. Oxford University Press, New York, 2nd edition, 1999.

12. E. K. Berndt, B. H. Hall, R. E. Hall, and J. A. Hausman. Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement*, 3:653–665, 1974.
13. C. R. Bhat. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research, Part B*, 35:677–693, 2001.
14. R. D. Bock. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37:29–51, 1972.
15. R. D. Bock. *Multivariate Statistical Methods in Behavioral Research*. McGraw-Hill, New York, 1975.
16. R. D. Bock and M. Aitkin. Marginal maximum likelihood estimation of item parameters: An application of the EM algorithm. *Psychometrika*, 46:443–459, 1981.
17. R. D. Bock and M. Lieberman. Fitting a response model for n dichotomously scored items. *Psychometrika*, 35:179–197, 1970.
18. R. D. Bock and S. Shilling. High-dimensional full-information item factor analysis. In M. Berkane, editor, *Latent Variable Modeling and Applications to Causality*, pages 163–176. Springer, New York, 1997.
19. N. E. Breslow and X. Lin. Bias correction in generalised linear mixed models with a single component of dispersion. *Biometrika*, 82:81–91, 1995.
20. G. Camilli and L. A. Shepard. *Methods for Identifying Biased Test Items*. Sage, Thousand Oaks, CA, 1994.
21. M. R. Conaway. Analysis of repeated categorical measurements with conditional likelihood methods. *Journal of the American Statistical Association*, 84: 53–61, 1989.
22. C. Corcoran, B. Coull, and A. Patel. *EGRET for Windows User Manual*. CYTEL Software Corporation, Cambridge, MA, 1999.
23. L. A. Cupples, R. B. D'Agostino, K. Anderson, and W. B. Kannel. Comparison of baseline and repeated measure covariate techniques in the Framingham Heart Study. *Statistics in Medicine*, 7:205–218, 1985.
24. R. B. D'Agostino, M.-L. Lee, A. J. Belanger, L. A. Cupples, K. Anderson, and W. B. Kannel. Relation of pooled logistic regression to time dependent Cox regression analysis: The Framingham Heart Study. *Statistics in Medicine*, 9: 1501–1515, 1990.
25. M. J. Daniels and C. Gatsonis. Hierarchical polytomous regression models with applications to health services research. *Statistics in Medicine*, 16:2311–2325, 1997.
26. K. A. Doksum and M. Gasko. On a correspondence between models in binary regression analysis and in survival analysis. *International Statistical Review*, 58:243–252, 1990.
27. D. M. Drukker. Maximum simulated likelihood: Introduction to a special issue. *The Stata Journal*, 6:153–155, 2006.
28. J. Engel. On the analysis of grouped extreme-value data with GLIM. *Applied Statistics*, 42:633–640, 1993.
29. F. Ezzet and J. Whitehead. A random effects model for ordinal responses from a crossover trial. *Statistics in Medicine*, 10:901–907, 1991.

30. A. Fielding, M. Yang, and H. Goldstein. Multilevel ordinal models for examination grades. *Statistical Modelling*, 3:127–153, 2003.
31. G. M. Fitzmaurice, N. M. Laird, and A. G. Rotnitzky. Regression models for discrete longitudinal responses. *Statistical Science*, 8:284–309, 1993.
32. S. A. Freels, R. B. Warnecke, T. P. Johnson, and B. R. Flay. Evaluation of the effects of a smoking cessation intervention using the multilevel thresholds of change model. *Evaluation Review*, 26:40–58, 2002.
33. S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.
34. R. D. Gibbons and R. D. Bock. Trend in correlated proportions. *Psychometrika*, 52:113–124, 1987.
35. R. D. Gibbons and D. Hedeker. Random-effects probit and logistic regression models for three-level data. *Biometrics*, 53:1527–1537, 1997.
36. G. Glasgow. Mixed logit models for multiparty elections. *Political Analysis*, 9: 116–136, 2001.
37. H. Goldstein. Nonlinear multilevel models, with an application to discrete response data. *Biometrika*, 78:45–51, 1991.
38. H. Goldstein. *Multilevel Statistical Models*. Edward Arnold, London, 3rd edition, 2003.
39. H. Goldstein and J. Rasbash. Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society, Series A*, 159: 505–513, 1996.
40. W. H. Greene. *LIMDEP Version 8.0 User's Manual*. Econometric Software, Plainview, NY, 4th edition, 2002.
41. P. Haan and A. Uhlendorff. Estimation of multinomial logit models with unobserved heterogeneity using maximum simulated likelihood. *The Stata Journal*, 6:229–245, 2006.
42. A. Han and J. A. Hausman. Flexible parametric estimation of duration and competing risk models. *Journal of Applied Econometrics*, 5:1–28, 1990.
43. J. Hartzel, A. Agresti, and B. Caffo. Multinomial logit random effects models. *Statistical Modelling*, 1:81–102, 2001.
44. D. A. Harville and R. W. Mee. A mixed-model procedure for analyzing ordered categorical data. *Biometrics*, 40:393–408, 1984.
45. P. J. Heagerty and S. L. Zeger. Marginalized multilevel models and likelihood inference. *Statisical Science*, 15:1–26, 2000. (with discussion)
46. D. Hedeker. MIXNO: A computer program for mixed-effects nominal logistic regression. *Journal of Statistical Software*, 4(5):1–92, 1999.
47. D. Hedeker. A mixed-effects multinomial logistic regression model. *Statistics in Medicine*, 21:1433–1446, 2003.
48. D. Hedeker and R. D. Gibbons. A random-effects ordinal regression model for multilevel analysis. *Biometrics*, 50:933–944, 1994.
49. D. Hedeker and R. D. Gibbons. MIXOR: A computer program for mixed-effects ordinal regression analysis. *Computer Methods and Programs in Biomedicine*, 49:157–176, 1996.

50. D. Hedeker and R. D. Gibbons. Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods*, 2:64–78, 1997.
51. D. Hedeker and R. D. Gibbons. *Longitudinal Data Analysis*. Wiley, New York, 2006.
52. D. Hedeker and R. J. Mermelstein. A multilevel thresholds of change model for analysis of stages of change data. *Multivariate Behavioral Research*, 33: 427–455, 1998.
53. D. Hedeker and R. J. Mermelstein. Analysis of longitudinal substance use outcomes using random-effects regression models. *Addiction*, 95(Supplement 3):S381–S394, 2000.
54. D. Hedeker, O. Siddiqui, and F. B. Hu. Random-effects regression analysis of correlated grouped-time survival data. *Statistical Methods in Medical Research*, 9:161–179, 2000.
55. R. L. Hough, S. Harmon, H. Tarke, S. Yamashiro, R. Quinlivan, P. Landau-Cox, M. S. Hurlburt, P. A. Wood, R. Milone, V. Renker, A. Crowell, and E. Morris. Supported independent housing: Implementation issues and solutions in the San Diego McKinney homeless demonstration research project. In W. R. Breakey and J. W. Thompson, editors, *Mentally Ill and Homeless: Special Programs for Special Needs*, pages 95–117. Harwood, New York, 1997.
56. M. S. Hurlburt, P. A. Wood, and R. L. Hough. Providing independent housing for the homeless mentally ill: A novel approach to evaluating long-term longitudinal housing patterns. *Journal of Community Psychology*, 24:291–310, 1996.
57. H. Ishwaran. Univariate and multirater ordinal cumulative link regression with covariate specific cutpoints. *Canadian Journal of Statistics*, 28:715–730, 2000.
58. J. Jansen. On the statistical analysis of ordinal data when extravariation is present. *Applied Statistics*, 39:75–84, 1990.
59. E. Läärä and J. N. S. Matthews. The equivalence of two models for ordinal data. *Biometrika*, 72:206–207, 1985.
60. N. M. Laird. Missing data in longitudinal studies. *Statistics in Medicine*, 7: 305–315, 1988.
61. E. Lesaffre and B. Spiessens. On the effect of the number of quadrature points in a logistic random-effects model: An example. *Applied Statistics*, 50:325–335, 2001.
62. J. K. Lindsey and P. Lambert. On the appropriateness of marginal models for repeated measurements in clinical trials. *Statistics in Medicine*, 17:447–469, 1998.
63. L. C. Liu and D. Hedeker. A mixed-effects regression model for longitudinal multivariate ordinal data. *Biometrics*, 62:261–268, 2006.
64. Q. Liu and D. A. Pierce. A note on Gauss-Hermite quadrature. *Biometrika*, 81:624–629, 1994.
65. J. S. Long. *Regression Models for Categorical and Limited Dependent Variables*. Sage, Thousand Oaks, CA, 1997.
66. F. M. Lord. *Applications of Item Response Theory to Practical Testing Problems*. Erlbaum, Hillside, NJ, 1980.

67. R. D. Luce. *Individual Choice Behavior.* Wiley, New York, 1959.
68. G. S. Maddala. *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge University Press, Cambridge, UK, 1983.
69. J. R. Magnus. *Linear Structures.* Charles Griffin, London, 1988.
70. E. C. Marshall and D. Spiegelhalter. Institutional performance. In A. H. Leyland and H. Goldstein, editors, *Multilevel Modelling of Health Statistics*, pages 127–142. Wiley, New York, 2001.
71. P. McCullagh. Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B*, 42:109–142, 1980. (with discussion)
72. P. McCullagh and J. A. Nelder. *Generalized Linear Models.* Chapman & Hall, London, 2nd edition, 1989.
73. D. McFadden. Conditional logit analysis of qualitative choice behavior. In P. Zarembka, editor, *Frontiers in Econometrics.* Academic Press, New York, 1973.
74. D. McFadden. Qualitative response models. In W. Hildenbrand, editor, *Advances in Econometrics*, pages 1–37. Cambridge University Press, Cambridge, UK, 1980.
75. J. M. Neuhaus, J. D. Kalbfleisch, and W. W. Hauck. A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review*, 59:25–35, 1991.
76. J. F. Pendergast, S. J. Gange, M. A. Newton, M. J. Lindstrom, M. Palta, and M. R. Fisher. A survey of methods for analyzing clustered binary response data. *International Statistical Review*, 64:89–118, 1996.
77. B. Peterson and F. E. Harrell. Partial proportional odds models for ordinal response variables. *Applied Statistics*, 39:205–217, 1990.
78. J. C. Pinheiro and D. M. Bates. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, 4:12–35, 1995.
79. R. L. Prentice and L. A. Gloeckler. Regression analysis of grouped survival data with application to breast cancer data. *Biometrics*, 34:57–67, 1978.
80. S. Rabe-Hesketh, A. Skrondal, and A. Pickles. Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, 2: 1–21, 2002.
81. S. Rabe-Hesketh, A. Skrondal, and A. Pickles. GLLAMM manual. Working Paper 160, U.C. Berkeley Division of Biostatistics, Berkeley, CA, 2004. (Downloadable from `http://www.bepress.com/ucbbiostat/paper160/`)
82. S. Rabe-Hesketh, A. Skrondal, and A. Pickles. Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics*, 128:301–323, 2005.
83. R. Raman and D. Hedeker. A mixed-effects regression model for three-level ordinal response data. *Statistics in Medicine*, 24:3331–3345, 2005.
84. J. Rasbash, F. Steele, W. J. Browne, and B. Prosser. *A User's Guide to MLwiN. Version 2.0.* Centre for Multilevel Modelling, University of Bristol, Bristol, UK, 2005.
85. S. W. Raudenbush and A. S. Bryk. *Hierarchical Linear Models: Applications and Data Analysis Methods.* Sage, Thousand Oaks, CA, 2nd edition, 2002.

86. S. W. Raudenbush, A. S. Bryk, Y. F. Cheong, and R. Congdon. *HLM 6: Hierarchical Linear and Nonlinear Modeling*. Scientific Software International, Chicago, 2004.
87. S. W. Raudenbush, M.-L. Yang, and M. Yosef. Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation. *Journal of Computational and Graphical Statistics*, 9: 141–157, 2000.
88. D. Revelt and K. Train. Mixed logit with repeated choices: Household's choices of appliance efficiency level. *Review of Economics and Statistics*, 80:647–657, 1998.
89. F. Rijmen, F. Tuerlinckx, P. De Boeck, and P. Kuppens. A nonlinear mixed model framework for item response theory. *Psychological Methods*, 8:185–205, 2003.
90. G. Rodríguez and N. Goldman. An assessment of estimation procedures for multilevel models with binary responses. *Journal of the Royal Statistical Society, Series A*, 158:73–89, 1995.
91. D. B. Rubin. Inference and missing data. *Biometrika*, 63:581–592, 1976. (with discussion)
92. F. Samejima. Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph No. 17*, 1969.
93. SAS/Stat. *SAS/Stat User's Guide, version 9.1.* SAS Institute, Cary, NC, 2004.
94. T. H. Scheike and T. K. Jensen. A discrete survival model with random effects: An application to time to pregnancy. *Biometrics*, 53:318–329, 1997.
95. J. D. Singer and J. B. Willett. It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational and Behavioral Statistics*, 18:155–195, 1993.
96. J. D. Singer and J. B. Willett. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford University Press, Oxford, UK, 2003.
97. A. Skrondal and S. Rabe-Hesketh. Multilevel logistic regression for polytomous data and rankings. *Psychometrika*, 68:267–287, 2003.
98. A. Skrondal and S. Rabe-Hesketh. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Chapman & Hall/CRC, Boca Raton, FL, 2004.
99. T. A. B. Snijders and R. J. Bosker. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Sage, Thousand Oaks, CA, 1999.
100. D. J. Spiegelhalter, A. Thomas, N. G. Best, and W. R. Gilks. *BUGS: Bayesian Inference Using Gibbs Sampling, Version 0.50*. Technical Report, MRC Biostatistics Unit, Cambridge, UK, 1995.
101. StataCorp. *Stata Statistical Software: Release 9*. Stata Corporation, College Station, TX, 2005.
102. S. Stern. Simulation-based estimation. *Journal of Economic Literature*, 35: 2006–2039, 1997.
103. R. Stiratelli, N. M. Laird, and J. H. Ware. Random-effects models for serial observations with binary response. *Biometrics*, 40:961–971, 1984.
104. A. H. Stroud and D. Secrest. *Gaussian Quadrature Formulas*. Prentice Hall, Englewood Cliffs, NJ, 1966.

105. M. A. Tanner. *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. Springer, New York, 3rd edition, 1996.
106. T. R. Ten Have. A mixed effects model for multivariate ordinal response data including correlated discrete failure times with ordinal responses. *Biometrics*, 52:473–491, 1996.
107. T. R. Ten Have, A. R. Kunselman, and L. A. Tran. A comparison of mixed effects logistic regression models for binary response data with two nested levels of clustering. *Statistics in Medicine*, 18:947–960, 1999.
108. T. R. Ten Have and D. H. Uttal. Subject-specific and population-averaged continuation ratio logit models for multiple discrete time survival profiles. *Applied Statistics*, 43:371–384, 1994.
109. J. Terza. Ordinal probit: A generalization. *Communications in Statistics: Theory and Methods*, 14:1–11, 1985.
110. D. Thissen and L. Steinberg. A taxonomy of item response models. *Psychometrika*, 51:567–577, 1986.
111. L. L. Thurstone. Psychophysical analysis. *American Journal of Psychology*, 38:368–389, 1927.
112. K. E. Train. *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge, UK, 2003.
113. G. Tutz and W. Hennevogl. Random effects in ordinal regression models. *Computational Statistics & Data Analysis*, 22:537–557, 1996.
114. M. A. Wakefield, F. J. Chaloupka, N. J. Kaufman, C. T. Orleans, D. C. Barker, and E. E. Ruel. Effect of restrictions on smoking at home, at school, and in public places on teenage smoking: Cross sectional study. *British Medical Journal*, 321:333–337, 2001.
115. G. Y. Wong and W. M. Mason. The hierarchical logistic regression model for multilevel analysis. *Journal of the American Statistical Association*, 80: 513–524, 1985.
116. B. D. Wright. Solving measurement problems with the Rasch model. *Journal of Educational Measurement*, 14:97–116, 1977.
117. H. Xie, G. McHugo, A. Sengupta, D. Hedeker, and R. Drake. An application of the thresholds of change model to the analysis of mental health data. *Mental Health Services Research*, 3:107–114, 2001.
118. S. L. Zeger, K.-Y. Liang, and P. S. Albert. Models for longitudinal data: A generalized estimating equation approach. *Biometrics*, 44:1049–1060, 1988.