



*The World's Largest Open Access Agricultural & Applied Economics Digital Library*

**This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.**

**Help ensure our sustainability.**

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

[aesearch@umn.edu](mailto:aesearch@umn.edu)

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

*No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.*

## Reliable estimation of generalized linear mixed models using adaptive quadrature

Sophia Rabe-Hesketh      Anders Skrondal  
Institute of Psychiatry      National Institute of  
King's College London      Public Health, Oslo  
Andrew Pickles  
School of Epidemiology and Health Science  
The University of Manchester

**Abstract.** Generalized linear mixed models or multilevel regression models have become increasingly popular. Several methods have been proposed for estimating such models. However, to date there is no single method that can be assumed to work well in all circumstances in terms of both parameter recovery and computational efficiency. Stata's `xt` commands for two-level generalized linear mixed models (e.g., `xtlogit`) employ Gauss–Hermite quadrature to evaluate and maximize the marginal log likelihood. The method generally works very well, and often better than common contenders such as MQL and PQL, but there are cases where quadrature performs poorly. Adaptive quadrature has been suggested to overcome these problems in the two-level case. We have recently implemented a multilevel version of this method in `gllamm`, a program that fits a large class of multilevel latent variable models including multilevel generalized linear mixed models. As far as we know, this is the first time that adaptive quadrature has been proposed for multilevel models. We show that adaptive quadrature works well in problems where ordinary quadrature fails. Furthermore, even when ordinary quadrature works, adaptive quadrature is often computationally more efficient since it requires fewer quadrature points to achieve the same precision.

**Keywords:** `st0005`, adaptive quadrature, `gllamm`, generalized linear mixed models, random-effects models, panel data, numerical integration, adaptive integration, multilevel models, clustered data

### 1 Introduction

Generalized linear models, see, for example, McCullagh and Nelder (1989), are regression models for different response types. The models are constructed by defining a linear predictor

$$\eta_i = x_i' \beta$$

specifying the functional relationship or link between this predictor and the expectation of the response and selecting a distribution for the response given its expectation. Here  $x_i$  are explanatory variables for unit  $i$  with fixed effects  $\beta$ .

Generalized linear *mixed* models, see, for example, Goldstein (1995) and McCulloch and Searle (2001), include both fixed and random effects in the linear predictor. The

models are used for grouped or clustered data where observations within a cluster cannot be assumed to be mutually independent. Examples are panel data with repeated observations on subjects or two-stage surveys where elementary units are nested in primary sampling units; for example, children in schools. In the simplest generalized linear mixed models, the dependence structure of clustered data is modeled by introducing a random intercept into the linear predictor. The random intercept is shared by all units in the same cluster and can be interpreted as cluster level unobserved heterogeneity. Let  $i$  index the elementary “level 1” units and  $j$  index the clusters or “level 2” units. Conditional on the random intercept, the model is a generalized linear model with linear predictor

$$\eta_{ij} = x'_{ij}\beta + u_j^{(2)} \quad (1)$$

where  $x_{ij}$  are explanatory variables with fixed coefficients  $\beta$ , and  $u_j^{(2)}$  is a random intercept at level 2. The random intercept is assumed to have a normal distribution with mean 0. Models of this kind can be fitted using Stata’s panel data commands, `xtreg`, `xtlogit`, `xtpois`, `xtclog`, and so on.

The model can be extended in two ways. First, random coefficients can be included, allowing the effects of covariates to vary between level-2 units. Second, if the data are multilevel with several levels of clustering (e.g., repeated observations on children in schools), higher level random effects (intercepts and coefficients) can be included in the model. Omitting subscripts for the units of observation, the most general form of the linear predictor for the  $L$ -level case is

$$\eta = x'\beta + \sum_{l=2}^L z^{(l)'}u^{(l)}$$

where  $z^{(l)}$  are covariates (generally subsets of  $x$ ) with random effects  $u^{(l)}$  at level  $l$ . The first element of each  $z^{(l)}$  is typically equal to 1 so that the first element of  $u^{(l)}$  represents a random intercept. The random effects at level  $l$  are assumed to have a multivariate normal “prior” distribution with

$$u^{(l)} \sim N(0, \Sigma_l)$$

and to be independent of the random effects at any of the other levels. These multilevel and random-coefficient models are not available among Stata’s `xt` panel data models but can be fitted using `gllamm`; see Rabe-Hesketh et al. (2000), Rabe-Hesketh et al. (2001a), and Rabe-Hesketh et al. (2001b), since they are special cases of “generalized linear latent and mixed models”.

In Section 2, we discuss estimation of generalized linear mixed models and describe in detail the implementation in `gllamm` of both quadrature and adaptive quadrature; in Section 3, we describe the syntax of `gllamm`; and in Section 4, we analyze two datasets that have previously been analyzed by other methods.

## 2 Estimation of generalized linear mixed models

The likelihood of the observed data is a marginal likelihood where the random effects have been “integrated out”. Unfortunately, this marginal likelihood does not generally have a closed-form expression and approximate methods of estimation must be used. The most commonly used methods include marginal quasi-likelihood (MQL), penalized quasi-likelihood (PQL), Markov Chain Monte Carlo (MCMC), and Gaussian quadrature (GQ). A promising improvement of Gaussian quadrature is adaptive Gaussian quadrature (AGQ).

### 2.1 MQL, PQL, and MCMC

The methods MQL and PQL have been derived in a number of ways; see, for example, McCulloch and Searle (2001). Here we summarize the description in Goldstein (1995). The idea is that linear models can be estimated in a straightforward way by iterating between (1) generalized least squares to estimate the fixed parameters  $\beta$  for a given covariance matrix of the responses (which depends on the  $\Sigma_l$ ), and (2) estimating the  $\Sigma_l$  from the residuals  $y - x'\beta$ , an algorithm known as iterative generalized least squares (IGLS). MQL and PQL are based on approximating generalized linear mixed models as linear mixed models so that the IGLS algorithm can be applied.

In generalized linear mixed models, the expectation of the response is

$$\mu = g^{-1}(x'\beta + \sum_{l=1}^L z^{(l)'}u^{(l)}) \quad (2)$$

where  $g^{-1}(\cdot)$  is an inverse link function. For given current parameter estimates, this model can be approximated as a linear mixed model by expanding the inverse link function as a Taylor series expansion and representing the variability of the response  $y$  around its mean by a heteroskedastic error term with variance function equal to that of the chosen distribution family. That is, for dichotomous responses,

$$y = \mu + \epsilon\sqrt{\pi(1-\pi)}$$

where  $\text{var}(\epsilon) = 1$  and  $\pi$  is the current predicted probability.

In MQL and PQL, (2) is expanded around the “current” linear predictor. A first-order Taylor series expansion is used with respect to the fixed effects and a first- or second-order expansion with respect to the random effects. In MQL, the current linear predictor is obtained by substituting current estimates for the fixed effects and zero for the random effects. In PQL, the expansion is improved by setting the random effects equal to their posterior means instead of zero. MQL and PQL are available in MLwiN (see Goldstein et al. 1998) and in HLM (see Bryk et al. 1996).

MCMC is a simulation approach. By assuming an a priori distribution for the model parameters, samples of parameter values are drawn from their posterior distribution using simulation. Since the required joint posterior distribution is generally intractable,

it is not possible to simulate directly from this distribution. Instead, a Markov chain is used in which subsets of parameters are sequentially sampled from their conditional distributions given current values of the other parameters. After a burn-in period, when this chain has become stationary, the sampled parameter values follow the required distribution. The mean, median, or mode of these sampled parameters are then used as parameter estimates. The parameter estimates are essentially maximum likelihood estimates if the prior distributions are vague or diffuse. MCMC is available in MLwin and in BUGS; see Spiegelhalter et al. (1996).

## 2.2 Gaussian quadrature

Gaussian quadrature can be used to approximate the marginal likelihood by numerical integration. Let  $\theta$  be the vector of all model parameters including the fixed coefficients  $\beta$  and the non-duplicated elements of the covariance matrices  $\Sigma_l$ . Further let  $y^{(l)}$  be the response vector,  $X^{(l)}$  the design matrix with rows  $(x', z^{(2)'} , \dots, z^{(L)'} )$  for all units in a particular level- $l$  unit and  $U^{(l)} = (u^{(l)'}, \dots, u^{(L)'} )'$ . Let the conditional likelihood contribution of a level-1 unit given the random effects be denoted  $f^{(1)}(\theta; y^{(1)}, X^{(1)}|U^{(2)})$ . Depending on the chosen family, this could, for instance, be a Poisson or binomial probability. The conditional likelihood contribution of a level-2 unit given the random effects at levels 3 and above is

$$f^{(2)}(\theta; y^{(2)}, X^{(2)}|U^{(3)}) = \int g^{(2)}(u^{(2)}) \prod f^{(1)}(\theta; y^{(1)}, X^{(1)}|U^{(2)}) du^{(2)}$$

where  $g^{(l)}(u^{(l)})$  is the multivariate normal density with mean 0 and covariance matrix  $\Sigma_l$  and the product is over all level-1 units within the level-2 unit. The conditional likelihood contribution of an  $l$ -level unit is obtained recursively from the conditional likelihood contributions of the  $(l-1)$ -level units within it.

$$f^{(l)}(\theta; y^{(l)}, X^{(l)}|U^{(l+1)}) = \int g^{(l)}(u^{(l)}) \prod f^{(l-1)}(\theta; y^{(l-1)}, X^{(l-1)}|U^{(l)}) du^{(l)} \quad (3)$$

The total likelihood is

$$f(\theta; y, X) = \prod \int g^{(L)}(u^{(L)}) \prod f^{(L-1)}(\theta; y^{(L-1)}, X^{(L-1)}|u^{(L)}) du^{(L)}$$

where  $y$  and  $X$  are the vector of responses and design matrix for all units and the first product is over all highest level units.

For given parameter values, the multivariate integral over the  $M_l$  random effects variables  $u^{(l)}$  at level  $l$  is evaluated by integrating over  $M_l$  independent standard normally distributed random effects  $v^{(l)}$  with  $u^{(l)} = C_l v^{(l)}$ , where  $C_l$  is the Cholesky decomposition of  $\Sigma_l$ , i.e.,  $\Sigma_l = C_l' C_l$ . Let  $V^{(l)} = (v^{(l)'}, \dots, v^{(L)'} )'$ . The integral can then be approximated by Cartesian product quadrature as

$$\begin{aligned}
& \int g^{(l)}(u^{(l)}) \prod f^{(l-1)}(\theta; y^{(l-1)}, X^{(l-1)} | U^{(l)}) du^{(l)} \\
&= \int \phi(v_{M_l}^{(l)}) \cdots \int \phi(v_1^{(l)}) \prod f^{(l-1)}(\theta; y^{(l-1)}, X^{(l-1)} | V^{(l)}) dv_1^{(l)} \cdots dv_{M_l}^{(l)} \\
&\approx \sum_{r_{M_l}=1}^R p_{r_{M_l}} \cdots \sum_{r_1=1}^R p_{r_1} \prod f^{(l-1)}(\theta; y^{(l-1)}, X^{(l-1)} | a_{r_1}, \dots, a_{r_{M_l}}, V^{(l)}) \quad (4)
\end{aligned}$$

where  $\phi(\cdot)$  is the standard normal density and  $p_r$  and  $a_r$  are the  $r$ th quadrature weight and location of  $R$ -point Gauss–Hermite quadrature. Gaussian quadrature is available for two-level models in Stata, MIXOR, see Hedeker and Gibbons (1996), and in SAS PROC NLMIXED, see Wolfinger (1999). For multilevel models, quadrature is available in aML, see Lillard and Panis (2000), and in `gllamm`.

### 2.3 Adaptive quadrature

Consider the quadrature approximation for the likelihood contribution of a level-2 unit  $j$  for the simple random intercept model in (1):

$$f^{(2)}(\theta; y_j^{(2)}, X_j^{(2)}) = \int \phi(v_j^{(2)}) \prod_i f^{(1)}(\theta; y_{ij}^{(1)}, X_{ij}^{(1)} | v_j^{(2)}) dv_j^{(2)} \quad (5)$$

$$\approx \sum_{r=1}^R p_r \prod_i f^{(1)}(\theta; y_{ij}^{(1)}, X_{ij}^{(1)} | a_r) \quad (6)$$

The approximation is exact if the product  $\prod_i f^{(1)}(\theta; y_{ij}^{(1)}, X_{ij}^{(1)} | v_j^{(2)})$  is a  $2R - 1$  degree polynomial in  $v_j^{(2)}$ . However, as pointed out by Albert and Follmann (2000) and Lesaffre and Spiessens (2001), this product often has a sharp peak and is poorly approximated by a low-degree polynomial. The peak may be located between adjacent quadrature locations so that a substantial contribution to the likelihood is lost. The product will tend to have a sharper peak if there is a larger number of level-1 units within the level-2 unit and if the individual densities have their peaks in similar locations (i.e., if there is a high intraclass correlation). In addition, in the case of Poisson data, the individual terms in the product will have sharper peaks if the counts are higher. Therefore, larger cluster sizes, larger numbers of events, and large intraclass correlations can have similar detrimental effects on the quadrature approximation.

Note that the integrand in equation (5) is the product of the prior density of  $v_j^{(2)}$  and the joint probability (density) of the responses given  $v_j^{(2)}$  which, after normalization with respect to  $v_j^{(2)}$ , is just the posterior density of  $v_j^{(2)}$  given the observed responses. For large cluster sizes, this posterior density will be approximately normal, see, for example, Carlin and Louis (1998, 142–144). Let  $\mu_j$  and  $\tau_j^2$  be the mean and variance of the posterior density. Then for large cluster sizes, the integrand is approximately

proportional to the normal density  $\phi(v_j^{(l)}; \mu_j, \tau_j^2)$  with mean  $\mu_j$  and variance  $\tau_j^2$ . Writing the integral as

$$f^{(2)}(\theta; y_j^{(2)}, X_j^{(2)}) = \int \phi(v_j^{(2)}; \mu_j, \tau_j^2) \left\{ \frac{\phi(v_j^{(2)}) \prod_i f^{(1)}(\theta; y_{ij}^{(1)}, X_{ij}^{(1)} | v_j^{(2)})}{\phi(v_j^{(2)}; \mu_j, \tau_j^2)} \right\} dv_j^{(2)} \quad (7)$$

changing the variable of integration from  $v_j^{(2)}$  to  $\zeta_j = (v_j^{(2)} - \mu_j)/\tau_j$ , and applying the standard quadrature rule gives

$$f_j^{(2)}(\theta; y_j^{(2)}, X_j^{(2)}) \approx \sum_{r=1}^R \pi_{jr} \prod_{i=1}^{n_j} f_{ij}^{(1)}(\theta; y_{ij}^{(1)}, X_{ij}^{(1)} | \alpha_{jr})$$

where

$$\alpha_{jr} = \mu_j + \tau_j a_r \quad (8)$$

$$\pi_{jr} = \sqrt{2\pi} \tau_j \exp(a_r^2/2) \phi(\mu_j + \tau_j a_r) p_r \quad (9)$$

This approximation assumes that the ratio in brackets in equation (7) is well approximated by a low degree polynomial, which will be the case if the numerator is approximately proportional to the denominator. We have therefore essentially approximated the posterior density by a normal density with the same mean and standard deviation. The posterior means and standard deviations can easily be computed along with the log likelihood as suggested for Bayesian inference by Naylor and Smith (1982), and this method is implemented in **gllamm**. An alternative is to approximate the posterior density by a normal density with the same mode and the same curvature at the mode; see Liu and Pierce (1994). This first-order Laplace approximation takes  $\mu_j$  to be the mode and  $\tau_j$  to be the negative inverse of the second derivative of the log posterior density at the mode. This approach is implemented for two-level models in SAS NL MIXED, see Wolfinger (1999), and for exploratory factor models for binary items in TESTFACT, see Bock and Schilling (1997) and Bock et al. (1999). The approaches coincide when the posterior density is normal. We use the approach by Naylor and Smith since it is much easier to generalize to multilevel models. Pinheiro and Bates (1995) point out that while ordinary quadrature is essentially a deterministic version of simple Monte Carlo integration, adaptive quadrature is a deterministic version of importance sampling with  $\phi(v_j^{(2)}; \mu_j, \tau_j^2)$  as importance density.

Figures 1 and 2 compare ordinary Gaussian quadrature (GQ) and adaptive Gaussian quadrature (AGQ) for a hypothetical cluster with normal prior and posterior densities. The bars represent the quadrature weights at the quadrature locations for 5-point GQ and AGQ, respectively. It is clear that unlike GQ, AGQ centers the locations under the peak of the integrand (proportional to the posterior density) and spreads them out appropriately.

(Continued on next page)

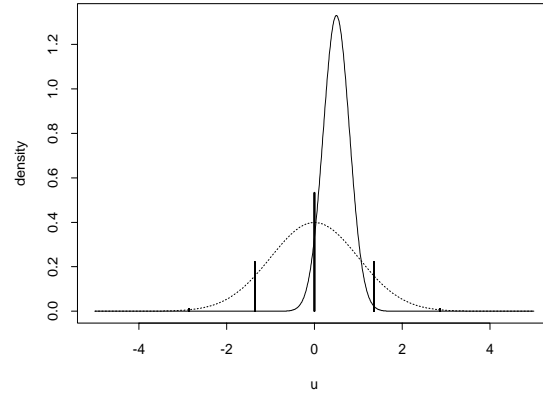


Figure 1: Prior (solid curve) and posterior (dotted curve) densities and quadrature weights (bars) for GA.

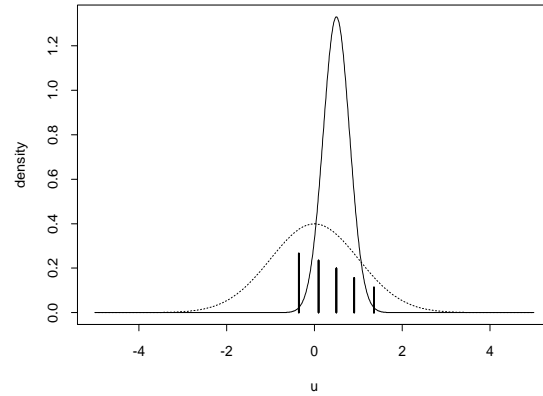


Figure 2: Prior (solid curve) and posterior (dotted curve) densities and quadrature weights (bars) for AGQ.

As shown in equation (3), the likelihood for general multilevel models is obtained by a recursive method where the likelihood contribution of a level- $l$  unit, conditional on higher level random effects  $V^{(l)}$ , is obtained by integrating out the random effects at level  $l$ . In equation (4), this is done by first integrating over  $v_1^{(l)}$ , then over  $v_2^{(l)}$ , up to  $v_M^{(l)}$ . To apply adaptive quadrature to each of these univariate integrals, we would have to use the posterior mean and standard deviation of each random effect *conditional* on all same-level and higher level random effects not yet integrated over. Since this would require extremely heavy computation, we transform the variables so that they have zero



posterior correlations as suggested by Naylor and Smith (1982). The *marginal* means and standard deviations of the orthogonalized random effects can then be used.

Since these posterior means  $\mu_m^{(l)}$  and standard deviations  $\tau_m^{(l)}$  depend on the parameter estimates, the algorithm implemented in `gllamm` alternates between predicting (in the  $k$ th iteration) the posterior means  $\mu_m^{(l)k}$  and standard deviations  $\tau_m^{(l)k}$  for given parameters  $\theta^{k-1}$  and updating the parameters to  $\theta^k$  using adaptive quadrature based on  $\mu_m^{(l)k}$  and  $\tau_m^{(l)k}$ . Naylor and Smith (1982) use a similar iterative method. The algorithm can be outlined as follows:

- Obtain starting values for the parameters  $\theta^0$
- Repeat for  $k = 1, 2, \dots$  until convergence:
  - Predict the posterior means and standard deviations  $\mu_m^{(l)k}$  and  $\tau_m^{(l)k}$ :
    - \* Predict the posterior means and standard deviations  $\mu_m^{(l)k0}$  and  $\tau_m^{(l)k0}$  using quadrature based on  $\mu_m^{(l)k-1}$  and  $\tau_m^{(l)k-1}$  (0 and 1 for  $k = 1$ ).
    - \* Repeat for  $j = 1, 2, \dots$  until convergence:
      - Predict the posterior means and standard deviations  $\mu_m^{(l)kj}$  and  $\tau_m^{(l)kj}$  using adaptive quadrature based on  $\mu_m^{(l)kj-1}$  and  $\tau_m^{(l)kj-1}$ .
  - Update the parameters to  $\theta^k$  using adaptive quadrature based on  $\mu_m^{(l)k}$  and  $\tau_m^{(l)k}$ .

Here the parameters are updated using the modified Newton–Raphson procedure implemented in Stata’s `ml maximize` (method `d0`) command. A non-stringent convergence criterion is used to determine the number of iterations of `ml maximize` the first time the parameters are updated ( $k = 1$ ), and a single iteration is used for  $k > 1$  except for the last iteration when the parameters are updated until convergence by the conventional convergence criteria used by `ml`.

One consequence of keeping the adaptive quadrature points fixed during the Newton–Raphson procedure is that we cannot use only one quadrature point since the log likelihood would be flat with respect to the covariance parameters  $\Sigma_l$ . In fact, in our experience, the method often requires five or more quadrature points to work. This is because the accuracy of posterior means and standard deviations for given parameter estimates is largely determined by the number of quadrature points. In particular, the predicted posterior standard deviations will be zero if only one of the quadrature locations makes a real contribution to the integral.

Our program uses numerical derivatives of the numerically integrated log-likelihood for the modified Newton–Raphson algorithm. While use of analytical derivatives would make the program considerably faster, the derivatives themselves would require numerical integration. Lesaffre and Spiessens (2001) show that the integrals required for the analytical derivatives are often more poorly approximated by quadrature and adaptive

quadrature than the integrals required for the marginal likelihood itself. It may therefore well be the case that numerical differentiation is more accurate than analytical differentiation.

## 2.4 A comparison of methods

MQL and PQL are computationally the most efficient of the methods described. These methods work well when the conditional distribution of the responses given the random effects is close to normal, for example, with a Poisson distribution if the mean is 7 or greater (see McCullagh and Nelder 1989), or if the responses are proportions with large binomial denominators. The method also works well if the conditional joint densities of the responses belonging to the clusters are nearly normal or, equivalently, if the posterior distribution of the random effects is nearly normal. Even for dichotomous responses, this increasingly becomes the case as the cluster sizes increase. However, both MQL and PQL perform poorly for dichotomous data with small cluster sizes; see, for example, Rodriguez and Goldman (1995), Breslow and Lin (1995), Lin and Breslow (1995), Goldstein and Rasbash (1996), and Browne and Draper (2002). In such situations, PQL is a better approximation than MQL, and second-order expansions of the random part (MQL-2 or PQL-2) yield better results than first-order expansions (MQL-1 or PQL-1).

MCMC appears to be a promising alternative to MQL/PQL. Browne and Draper (2002) showed that it performs better than PQL-2 in a simulation study of dichotomous three-level data (see Section 4). Another advantage of MCMC is that the availability of the sampled parameter values allows the properties of arbitrary functions of the parameters to be examined. However, the method is computationally demanding. Further, it can be difficult to determine when the chain has reached a stationary distribution and the method does not provide an empirical check of identification (see Keane 1992); an important consideration in more complex multilevel latent variable models.

Gaussian quadrature tends to work well if the responses are dichotomous and the cluster sizes are small to moderate (see [R] **quadchk** and [R] **xtlogit**), precisely in those situations where MQL/PQL tends to fail; see Rodriguez and Goldman (2001), Rabe-Hesketh et al. (2001c), and Stryhn et al. (2000). However, a large number of quadrature points are often needed to obtain a close approximation to the likelihood; see, for example, Crouch and Spiegelman (1990). Consequently, the methods can be computationally intensive, particularly if there are a large number of random effects. More importantly, Gaussian quadrature can fail even for binary data with small cluster sizes if the intraclass correlation is very high; see, for example, Lesaffre and Spiessens (2001). The approximation can also be poor if the responses are conditionally Poisson distributed; see, for example, Albert and Follmann (2000). In contrast to MQL/PQL, the performance of quadrature can easily be assessed by comparing solutions with different numbers of quadrature points (see [R] **quadchk**).

There are two versions of adaptive quadrature that differ in the choice of location and scale parameters  $\mu_m^{(l)}$  and  $\tau_m^{(l)}$ . AGQ-a sets the location  $\mu_m^{(l)}$  equal to the mode of the integrand and the scale  $\tau_m^{(l)}$  equal to the standard deviations of the normal density

approximating the integrand at the mode. When only one quadrature point is used, AGQ-a is equivalent to using a first order Laplace approximation, which is exact for linear mixed models; see Pinheiro and Bates (1995). AGQ-b instead uses the posterior means and standard deviations for  $\mu_m^{(l)}$  and  $\tau_m^{(l)}$ , which is equivalent to AGQ-a for linear mixed models except that  $\mu_m^{(l)}$  and  $\tau_m^{(l)}$  are themselves approximated by adaptive quadrature.

Adaptive quadrature is also expected to work well for other generalized linear mixed models. First, it should work well if the posterior densities are nearly normal (e.g., large cluster sizes and/or counts). Second, it should work well when the posterior densities are highly nonnormal and not too peaked (e.g., dichotomous responses, small cluster sizes and moderate intraclass correlation) since ordinary quadrature works well in this case. Therefore, the exact positioning of the quadrature locations is not crucial. When the posterior densities are highly non-normal but with sharp peaks (e.g., dichotomous responses, small cluster sizes and large intraclass correlation), it is not clear whether AGQ-a or AGQ-b is superior. While AGQ-a will capture the peak itself, the scale  $\tau_m^{(l)}$  may be too small to capture the remainder of the integrand since the approximating normal density at the mode will fall off too sharply. AGQ-b, on the other hand, may miss the exact peak, but the scale will be larger so that important contributions in the neighborhood of the peak may be captured.

We are not aware of any empirical studies comparing adaptive quadrature with other methods for generalized linear mixed models. Here we make such comparisons for two examples. In the first example, the responses are large counts, and PQL is expected to work well; whereas ordinary quadrature is expected to perform poorly. In the second example, the data are dichotomous and PQL has been shown to be inadequate. In both cases, adaptive quadrature is expected to perform better than ordinary quadrature in terms of parameter recovery and computational efficiency.

### 3 Syntax

The program `gllamm` can fit a large class of “generalized linear latent and mixed models”. Here we will confine our discussion to generalized linear mixed models. The syntax required for such models (omitting many available options) is

```
gllamm depvar [varlist] [if exp] [in range] , i(varlist) [ nrf(numlist)
    eqs(eqnames) nip(numlist) adapt family(string) link(string) noconstant
    offset(varname) eform ]
```

(Continued on next page)

families	links
<u>gaussian</u>	<u>identity</u>
<u>poisson</u>	log
<u>gamma</u>	<u>reciprocal</u>
<u>binomial</u>	<u>logit</u>
	<u>probit</u>
	c11 (complimentary log-log)
	<u>ologit</u> (o stands for ordinal)
	<u>oprobit</u>
	<u>ocl1</u>
	<u>mlogit</u>
	<u>sprobit</u> (scaled probit)

### 3.1 Options

**i**(*varlist*) gives the variables that define the hierarchical, nested clusters, from the lowest level (finest clusters) to the highest level; e.g., **i**(pupil class school). This option is required.

**nrf**(*numlist*) specifies the number of random effects at each level of clustering; i.e., for each variable in **i**(*varlist*). The default is **nrf**(1,...,1).

**eqs**(*eqnames*) specifies the equation names (defined before running **gllamm**) for the variables multiplying the random effects. The equations for the level-2 random effects are listed first, followed by those for the level-3 random effects, etc., with the number of equations per level specified in the **nrf**() option. If required, constants should be explicitly included in the equation definitions using variables equal to 1. If the option is not used, the random effects are assumed to be random intercepts, and only one random effect is allowed per level.

**nip**(*numlist*) specifies the number of quadrature points to be used for each random effect. If only one argument is given, the same number will be used for each random effect, the default being 8.

**adapt** specifies adaptive quadrature. The default is nonadaptive quadrature.

**family**(*string*), **link**(*string*), **noconstant**, **offset**(*varname*), and **eform** are defined in **glm** (see [R] **glm**).

See the **gllamm** manual (which is available along with the program from <http://www.iop.kcl.ac.uk/iop/Departments/BioComp/programs/gllamm.html>) for more options and examples.

## 4 Examples

### 4.1 Poisson regression example

We now consider the famous longitudinal epilepsy data from Thall and Vail (1990), also analyzed by Breslow and Clayton (1993). The data come from a randomized controlled trial comparing a new drug (`treat=1`) with placebo (`treat=0`). The outcomes are counts of epileptic seizures during the two weeks before each of four clinic visits (`visit`, coded  $-0.3, -0.1, 0.1, 0.3$ ). Breslow and Clayton used the log of a quarter of the number of seizures (`y`) in the eight weeks preceding entry into the trial (`lbas`) and the logarithm of age (`lage`) as covariates, in addition to a dummy variable for the fourth visit (`v4`) to account for a drop in seizure counts during the fourth interval. An interaction between `lbas` and `treat` (`lbas_trt`) was also included. Here we have subtracted the means of the predictors so that the model constant is not comparable with that in Breslow and Clayton.

Model II in Breslow and Clayton is a log-linear (Poisson regression) model with predictors `lbas`, `treat`, `lbas_trt`, `lage`, and `v4` and with a random intercept for subjects. The seizure count  $y_{ij}$  for patient  $j$  and visit  $i$  is assumed to be conditionally Poisson distributed with mean  $\mu_{ij}$ ,

$$\log(\mu_{ij}) = x'_{ij}\beta + u_{0j}^{(2)}$$

The subject-specific random intercept  $u_{0j}^{(2)}$  is assumed to have a normal distribution. The parameter estimates using PQL-1 are given in the first column of Table 1.

Table 1: Parameter estimates and standard errors for Models II and IV using PQL-1, see Breslow and Clayton (1993), and adaptive quadrature.

	Model II		Model IV	
	PQL-1	AGQ (10 points)	PQL-1	AGQ (7 points)
<i>Fixed effects</i>				
<code>lbas</code>	0.87 (0.14)	0.88 (0.13)	0.87 (0.14)	0.88 (0.13)
<code>treat</code>	-0.91 (0.41)	-0.93 (0.40)	-0.91 (0.41)	-0.93 (0.40)
<code>lbas_trt</code>	0.33 (0.21)	0.34 (0.20)	0.33 (0.21)	0.34 (0.20)
<code>lage</code>	0.47 (0.35)	0.48 (0.35)	0.46 (0.36)	0.48 (0.35)
<code>v4</code>	-0.16 (0.05)	-0.16 (0.05)		
<code>visit</code>			-0.26 (0.16)	-0.27 (0.16)
<i>Random effects</i>				
SD of intercept	0.53 (0.06)	0.50 (0.06)	0.52 (0.06)	0.50 (0.06)
SD of slope for visit			0.74 (0.16)	0.73 (0.16)
covariance			-0.01 (0.03)	0.00 (0.09)

#### Using non-adaptive quadrature

Model II is a two-level random intercept model and can be fitted using `xtpois`. We initially use 10-point quadrature:

```
. xtpois y lbas treat lbas_trt lage v4, i(subj) normal nolog quad(10)
Random-effects poisson                               Number of obs   =       236
Group variable (i) : subj                           Number of groups  =        59
Random effects u_i ~ Gaussian                       Obs per group: min =         4
                                                    avg =        4.0
                                                    max =         4
LR chi2(5) = 108.82
Prob > chi2 = 0.0000
Log likelihood = -666.02733
```

y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lbas	1.116892	.0739454	15.10	0.000	.9719621	1.261823
treat	-.8948228	.2752803	-3.25	0.001	-1.434362	-.3552834
lbas_trt	.3218361	.0977863	3.29	0.001	.1301785	.5134938
lage	.5565262	.2038638	2.73	0.006	.1569604	.956092
v4	-.1610871	.0545758	-2.95	0.003	-.2680537	-.0541205
_cons	2.15001	.1539091	13.97	0.000	1.848354	2.451666
/lnsig2u	-1.439468	.1455344	-9.89	0.000	-1.72471	-1.154226
sigma_u	.4868818	.035429			.4221668	.5615173
rho	.1916278	.0225442			.1512655	.2397181

Likelihood ratio test of rho=0:   chibar2(01) =   303.26 Prob>=chibar2 = 0.000

The parameter estimates do not agree very well with those of Table 1. To improve the quadrature approximation, we run the same model with 20-point quadrature:

```
. xtpois y lbas treat lbas_trt lage v4, i(subj) normal nolog quad(20)
Random-effects poisson                               Number of obs   =       236
Group variable (i) : subj                           Number of groups  =        59
Random effects u_i ~ Gaussian                       Obs per group: min =         4
                                                    avg =        4.0
                                                    max =         4
LR chi2(4) = 88.49
Prob > chi2 = 0.0000
Log likelihood = -666.97139
```

y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lbas	.9495236	.	.	.	.	.
treat	-1.425517	.2235628	-6.38	0.000	-1.863692	-.9873416
lbas_trt	.5863269	.1101971	5.32	0.000	.3703446	.8023092
lage	.697401	.2968941	2.35	0.019	.1154993	1.279303
v4	-.1599219	.0543964	-2.94	0.003	-.266537	-.0533069
_cons	2.392076	.	.	.	.	.
/lnsig2u	-1.344066	.2017095	-6.66	0.000	-1.739409	-.9487225
sigma_u	.5106694	.0515034			.4190753	.6222824
rho	.2068422	.0330922			.149388	.2791418

Likelihood ratio test of rho=0:   chibar2(01) =   301.38 Prob>=chibar2 = 0.000

The parameter estimates have changed considerably but are no closer to the PQL estimates. The large changes in parameter estimates together with the missing standard errors suggest that there is a problem with the quadrature approximation as confirmed by using `quadchk`:

```
. quadchk, nooutput
Refitting model quad() = 16
Refitting model quad() = 24
```

Quadrature check				
	Fitted quadrature 20 points	Comparison quadrature 16 points	Comparison quadrature 24 points	
Log likelihood	-666.97139	-664.70167 2.2697183 -.00340302	-665.5398 1.4315875 -.0021464	Difference Relative difference
y: lbas	.94952363	.87595585 -.07356778 -.07747862	.86799166 -.08153196 -.08586618	Difference Relative difference
y: treat	-1.4255166	-1.3410662 .08445034 -.05924192	-1.4585471 -.03303056 .02317094	Difference Relative difference
y: lbas_trt	.58632689	.60813569 .0218088 .03719563	.64108902 .05476212 .09339862	Difference Relative difference
y: lage	.69740102	.68233866 -.01506236 -.02159785	.76006521 .06266419 .08985389	Difference Relative difference
y: v4	-.15992194	-.16108712 -.00116518 .00728593	-.16108712 -.00116518 .00728593	Difference Relative difference
y: _cons	2.3920763	2.3389251 -.05315122 -.0222197	2.4200047 .0279284 .01167538	Difference Relative difference
lnsig2u: _cons	-1.3440659	-1.345403 -.00133714 .00099485	-1.1032861 .24077974 -.17914281	Difference Relative difference

The relative differences are large indicating that Gaussian quadrature is unreliable.

#### *Using adaptive quadrature*

We will now estimate the parameters of Model II using `gllamm` with 10-point adaptive quadrature. The syntax is as for `xtpois` except that the Poisson family is specified with the canonical log link as the default, `nip()` is used instead of `quad()`, and the `adapt` option is used:

```
. gllamm y lbas treat lbas_trt lage v4, i(subj) fam(poisson) nip(10) adapt
number of level 1 units = 236
number of level 2 units = 59
Condition Number = 9.3176111
gllamm model
log likelihood = -665.29073
```

y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lbas	.8844321	.1312308	6.74	0.000	.6272245	1.14164
treat	-.9330387	.4008309	-2.33	0.020	-1.718653	-.1474245
lbas_trt	.3382607	.2033363	1.66	0.096	-.0602711	.7367925
lage	.484237	.347276	1.39	0.163	-.1964114	1.164885
v4	-.1610871	.0545758	-2.95	0.003	-.2680537	-.0541206
_cons	2.114303	.2197154	9.62	0.000	1.683668	2.544937

Variances and covariances of random effects

```
***level 2 (subj)
var(1): .25282688 (.05895623)
```

The standard deviation of the random effect is estimated as  $\sqrt{.25282688} = 0.50$ . The standard error can be obtained using the delta method as  $.05895623/(\sqrt{2 \times .25282688}) = 0.06$ . The parameter estimates, also shown in column 2 of Table 1, agree closely with those of Breslow and Clayton.

Breslow and Clayton also considered a random coefficient model (Model IV) in which the effect of `visit`, denoted  $z_{ij}$ , varies randomly between subjects:

$$\log(\mu_{ij}) = x'_{ij}\beta + u_{0j}^{(2)} + u_{1j}^{(2)} z_{ij}$$

The subject specific random intercept  $u_{0j}^{(2)}$  and slope  $u_{1j}^{(2)}$  have a bivariate normal distribution. The fixed part of this model is the same as that of Model II except that the variable `visit` is used instead of `v4`.

To fit this in `gllamm`, we need to define equations for the variables with random coefficients, including the constant for the random intercept, and specify them using the `eqs()` option.

```
. gen cons=1
. eq subj: cons
. eq subj_v: visit
. gllamm y lbas treat lbas_trt lage visit, i(subj) fam(poisson) nrf(2)/*
>      */ eqs(subj subj_v) nip(7) adapt
number of level 1 units = 236
number of level 2 units = 59
Condition Number = 9.5770804
```



```
gllamm model
log likelihood = -655.68101
```

y	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lbas	.8849558	.1314142	6.73	0.000	.6273888	1.142523
treat	-.9295086	.3928139	-2.37	0.018	-1.69941	-.1596075
lbas_trt	.3384994	.1974394	1.71	0.086	-.0484747	.7254735
lage	.4767799	.3536353	1.35	0.178	-.2163326	1.169892
visit	-.2664214	.1646942	-1.62	0.106	-.5892161	.0563734
_cons	2.100037	.2147934	9.78	0.000	1.67905	2.521025

Variances and covariances of random effects

```
***level 2 (subj)
var(1): .25162631 (.05813468)
cov(1,2): .00289385 (.08872316) cor(1,2): .0079133
var(2): .5314739 (.22938506)
```

The standard deviations are  $\sqrt{0.25162631} = 0.50$  for the random intercept and  $\sqrt{0.5314739} = 0.73$  for the random slope, respectively. The delta method was used to obtain the standard errors. As shown in Table 1, the parameter estimates agree closely with those of Breslow and Clayton. For comparison, using non-adaptive quadrature with 8 points per dimension (not shown) produces baseline and treatment parameter estimates of 1.04 and  $-0.85$ , whereas 20 points per dimension gives 0.90 and  $-1.21$ , respectively.

## 4.2 Logistic regression example

Rodriguez and Goldman (1995) simulated data to produce the same structure as data from the 1987 Guatemalan National Survey of Maternal and Child Health. The outcome of interest, whether the women received prenatal care, was simulated for 2,449 births (level 1) by 1,558 women (level 2) from 161 communities (level 3). A logistic regression model with one covariate at each level and random intercepts at levels 2 and 3 was used

$$\eta_{ijk} = \beta_0 + \beta_1 x_{1ijk} + \beta_2 x_{2jk} + \beta_3 x_{3k} + u_{jk}^{(2)} + u_k^{(3)}$$

where  $i$  indexes the births,  $j$  the mothers, and  $k$  the communities.

The first of the simulated datasets (available from <http://www.blackwellpublishers.co.uk/rss/Readmefiles/goldman.htm>) has been analyzed by Browne and Draper (2002) using MQL-1, PQL-2, and MCMC with a diffuse prior. Their parameter estimates are given in Table 2 where it is clear that the PQL-2 parameter estimates are better than the MQL-1 estimates, as expected, but both methods underestimate the variances considerably.

Table 2: Parameter estimates for data simulated by Rodriguez and Goldman (1995)

	True	MQL-1	PQL-2	MCMC	GQ (10 points)	AGQ (5 points)
$\beta_0$	0.65	0.491 (0.149)	0.641 (0.186)	0.675 (0.209)	0.688 (0.207)	0.673 (0.202)
$\beta_1$	1.0	0.791 (0.172)	0.993 (0.201)	1.050 (0.225)	1.042 (0.221)	1.047 (0.221)
$\beta_2$	1.0	0.631 (0.081)	0.795 (0.099)	0.843 (0.115)	0.834 (0.112)	0.839 (0.112)
$\beta_3$	1.0	0.806 (0.189)	1.06 (0.237)	1.124 (0.268)	1.127 (0.260)	1.120 (0.260)
$\sigma_2^2$	1.0	0.000 (-)	0.486 (0.145)	0.921 (0.331)	0.886 (0.288)	0.881 (0.286)
$\sigma_3^2$	1.0	0.546 (0.102)	0.883 (0.159)	1.043 (0.217)	0.974 (0.197)	0.990 (0.203)

We can fit the three-level random intercept model in `gllamm` by specifying the level 2 and 3 clustering variables, `mother` and `comm` in the `i()` option and the logit link and binomial family in the `link()` and `family()` options, respectively. We will first use 10-point quadrature per dimension:

```
. gllamm care x1 x2 x3, i(mother comm) link(logit) family(binom) nip(10)
number of level 1 units = 2449
number of level 2 units = 1558
number of level 3 units = 161
Condition Number = 4.1067459
gllamm model
log likelihood = -1414.064
```

care	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
x1	1.042056	.221363	4.71	0.000	.6081927	1.47592
x2	.8335885	.1122263	7.43	0.000	.613629	1.053548
x3	1.127113	.2596609	4.34	0.000	.6181868	1.636039
_cons	.6881888	.2067724	3.33	0.001	.2829223	1.093455

#### Variances and covariances of random effects

```
***level 2 (mother)
var(1): .88572327 (.28812319)
***level 3 (comm)
var(1): .9736015 (.19671434)
```

The parameter estimates, in particular the variance estimates, are closer to the true values than the corresponding PQL-2 estimates. Increasing the number of quadrature points to 20 requires  $20 \times 20 = 400$  evaluations of the integrand and is relatively slow.

We may be able to achieve good accuracy with a lower number of quadrature points using adaptive quadrature. With 5 points (requiring  $5 \times 5 = 25$  evaluations) we get

```
. gllamm care x1 x2 x3, i(mother comm) link(logit) family(binom) nip(5) adapt
number of level 1 units = 2449
number of level 2 units = 1558
number of level 3 units = 161
Condition Number = 4.0249554
gllamm model
log likelihood = -1413.9554
```

care	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
x1	1.04719	.2211608	4.73	0.000	.613723	1.480657
x2	.8386616	.1116788	7.51	0.000	.6197751	1.057548
x3	1.120168	.2597512	4.31	0.000	.6110646	1.629271
_cons	.6726168	.2021648	3.33	0.001	.276381	1.068852

Variances and covariances of random effects

```
***level 2 (mother)
var(1): .8807801 (.28636287)
***level 3 (comm)
var(1): .98965411 (.20299419)
```

The estimates are very similar to the 10-point non-adaptive quadrature estimates. Increasing the number of quadrature points to 30 per dimension for non-adaptive quadrature and to 11 per dimension for adaptive quadrature gives essentially the same results. Therefore, we believe that the estimates are reliable.

## 5 Discussion

As far as we know, this is the first implementation of adaptive quadrature for multilevel generalized linear mixed models. Adaptive quadrature appears to be suitable when the posterior distribution is close to normal and when it is highly nonnormal, whereas ordinary quadrature fails in the first situation and PQL fails in the second. Adaptive quadrature is computationally more efficient than ordinary quadrature and other computer intensive methods such as MCMC. Another advantage of adaptive quadrature is that it provides a value for the maximized log likelihood useful for example for likelihood-ratio tests. In contrast to ordinary quadrature, adaptive quadrature also appears to give good parameter estimates for linear models, and although computationally less efficient than other methods such as IGLS, this will be useful for complex multilevel latent variable models that cannot yet be handled by other software. Extensive simulation studies are required to assess the performance of adaptive quadrature in a wide range of situations.

Although we have only discussed adaptive quadrature in the context of estimating generalized linear mixed models, `gllamm` allows this method to be used for estimat-

ing general multilevel latent variable models, including multilevel factor models and multilevel structural equation models; see Rabe-Hesketh et al. (2002). In addition to counts and dichotomous outcomes, `gllamm` can handle continuous, censored, ordinal, and nominal responses and rankings, see Skrondal and Rabe-Hesketh (2002), as well as continuous and discrete time survival data, see Rabe-Hesketh et al. (2001d).

A problem with `gllamm` is that it can be very slow, particularly if the models include many random effects. This is partly because `gllamm` is written in ado code, which needs to be interpreted by Stata while the program is running. Fortunately, Stata Corporation is converting parts of `gllamm` to internal code, which should result in a considerable increase in speed.

## 6 References

- Albert, P. S. and D. A. Follmann. 2000. Modeling repeated count data subject to informative dropout. *Biometrics* 56: 667–677.
- Bock, R. D., R. Gibbons, S. G. Shilling, E. Muraki, D. T. Wilson, and R. Wood. 1999. *TESTFACT: Test Scoring, Item Statistics, and Full Information Item Factor Analysis*. Chicago, IL: Scientific Software International.
- Bock, R. D. and S. Schilling. 1997. High-dimensional full-information item factor analysis. In *Latent Variable Modelling and Applications to Causality*, ed. M. Berkane, 164–176. New York: Springer-Verlag.
- Breslow, N. E. and D. G. Clayton. 1993. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 88: 9–25.
- Breslow, N. E. and X. Lin. 1995. Bias correction in generalised linear mixed models with a single component of overdispersion. *Biometrika* 82: 81–91.
- Browne, W. J. and D. Draper. 2002. A comparison of Bayesian and likelihood methods for fitting multilevel models. Submitted.
- Bryk, A. S., S. W. Raudenbush, and R. Congdon. 1996. *HLM. Hierarchical Linear and Nonlinear Modeling with the HLM/2L and HLM/3L Programs*. Chicago: Scientific Software International.
- Carlin, B. P. and T. A. Louis. 1998. *Bayes and Empirical Bayes Methods for Data Analysis*. Boca Raton: Chapman & Hall/CRC.
- Crouch, E. A. C. and D. Spiegelman. 1990. The evaluation of integrals of the form  $\int f(t) \exp(-t^2) dt$ : Application to logistic-normal models. *Journal of the American Statistical Association* 85: 464–469.
- Goldstein, H. 1995. *Multilevel Statistical Models*. London: Arnold.
- Goldstein, H. and J. Rasbash. 1996. Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society, Series A* 159: 505–513.

- Goldstein, H., J. Rasbash, I. Plewis, D. Draper, W. Browne, M. Yang, G. Woodhouse, and M. Healy. 1998. *A User's Guide to MLwiN*. London: Multilevel Models Project, Institute of Education, University of London.
- Hedeker, D. and R. D. Gibbons. 1996. MIXOR: a computer program for mixed-effects ordinal probit and logistic regression analysis. *Computer Methods and Programs in Biomedicine* 49: 157–176.
- Keane, M. P. 1992. A note on identification in the multinomial probit model. *Journal of Business and Economic Statistics* 10: 193–200.
- Lesaffre, E. and B. Spiessens. 2001. On the effect of the number of quadrature points in a logistic random-effects model: an example. *Applied Statistics* 50: 325–335.
- Lillard, L. A. and C. W. A. Panis. 2000. *Multiprocess Multilevel Modelling, aML Release 1, User's Guide and Reference Manual*. Los Angeles: EconoWare.
- Lin, X. and N. E. Breslow. 1995. Analysis of correlated binomial data in logistic-normal models. *Journal of Statistical Computing and Simulation* 55: 133–146.
- Liu, Q. and D. A. Pierce. 1994. A note on Gauss-Hermite quadrature. *Biometrika* 81: 624–629.
- McCullagh, P. and J. A. Nelder. 1989. *Generalized Linear Models*. 2d ed. London: Chapman and Hall.
- McCulloch, C. E. and A. F. M. Searle. 2001. *Generalized, Linear and Mixed Models*. New York: John Wiley & Sons.
- Naylor, J. C. and A. F. M. Smith. 1982. Applications of a method for the efficient computation of posterior distributions. *Applied Statistics* 31: 214–225.
- Pinheiro, J. C. and D. M. Bates. 1995. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational Graphics and Statistics* 4: 12–35.
- Rabe-Hesketh, S., A. Pickles, and A. Skrondal. 2001a. GLLAMM: A general class of multilevel models and a Stata program. *Multilevel Modelling Newsletter* 13: 17–23.
- . 2001b. *GLLAMM Manual*. London: Tech. Rep. 2001/01, Department of Biostatistics and Computing, Institute of Psychiatry, King's College, University of London. <http://www.iop.kcl.ac.uk/iop/departments/biocomp/programs/gllamm.html>
- Rabe-Hesketh, S., A. Pickles, and C. Taylor. 2000. sg129: Generalized linear latent and mixed models. *Stata Technical Bulletin* 53: 47–57. In *Stata Technical Bulletin Reprints*, vol. 9, 293–307. College Station, TX: Stata Press.
- Rabe-Hesketh, S., A. Skrondal, and A. Pickles. 2002. Generalized multilevel structural equation modelling. Submitted.

- Rabe-Hesketh, S., T. Touloupoulou, and R. M. Murray. 2001c. Multilevel modeling of cognitive function in schizophrenics and their first degree relatives. *Multivariate Behavioral Research*. 36: 279–298.
- Rabe-Hesketh, S., S. Yang, and A. Pickles. 2001d. Multilevel models for censored and latent responses. *Statistical Methods in Medical Research*. 10: 409–427.
- Rodriguez, B. and N. Goldman. 1995. An assessment of estimation procedures for multilevel models with binary responses. *Journal of the Royal Statistical Society, Series A* 158: 73–89.
- . 2001. Improved estimation procedures for multilevel models with binary response: a case study. *Journal of the Royal Statistical Society, Series A* 164: 339–355.
- Skrondal, A. and S. Rabe-Hesketh. 2002. Multilevel logistic regression for polytomous data and rankings. Submitted.
- Spiegelhalter, D., A. Thomas, N. Best, and W. Gilks. 1996. *BUGS 0.5 Bayesian Analysis using Gibbs sampling. Manual (version ii)*. London: Cambridge: MRC-Biostatistics Unit.  
<http://www.mrc-bsu.cam.ac.uk/bugs/documentation/contents.shtml>
- Stryhn, H., I. Dohoo, E. Tillard, and T. Hagedorn-Olsen. 2000. Simulation as a tool of validation in hierarchical generalized linear models. In *International Symposium on Veterinary Epidemiology and Economics*, 1136–1138. Beckenridge, CO.
- Thall, P. F. and S. C. Vail. 1990. Some covariance models for longitudinal count data with overdispersion. *Biometrics* 46: 657–671.
- Wolfinger, R. D. 1999. Fitting nonlinear mixed models with the new NLMIXED procedure. In *Proceedings of the 24th SAS User's Group Meeting*. Cary, NC: SAS Institute, Inc.

#### About the Authors

Sophia Rabe-Hesketh is Senior Lecturer in Statistics at the Department of Biostatistics and Computing, Institute of Psychiatry, King's College London.

Anders Skrondal is Senior Biostatistician at the Department of Epidemiology, National Institute of Public Health, Oslo.

Andrew Pickles is Professor of Epidemiological and Social Statistics at the School of Epidemiology and Health Science and the Cathy Marsh Centre for Census and Survey Research, The University of Manchester.