Dipartimento di Statistica
"Giuseppe Parenti"

Dipartimento di Statistica "G. Parenti" – Viale Morgagni 59 – 50134 Firenze - www.ds.unifi.it

QUADERNI 2001/48

# Alternative Simulation-Based Estimators of Logit Models with Random Effects

Giorgio Calzolari, Fabrizia Mealli, Carla Rampichini

Università degli Studi di Firenze

# Alternative Simulation-Based Estimators of Logit Models with Random Effects

Giorgio Calzolari, Fabrizia Mealli, Carla Rampichini
Università di Firenze, Dipartimento di Statistica "G. Parenti"

Quaderni del Dipartimento di Statistica "G. Parenti"

**SUMMARY**

Logit models with random effects are now widely used in applied Statistics and Econometrics. They usually lead to intractable likelihood functions, as they involve integrals without closed form solution. Numerical integration can be used to compute the likelihood and software is available (Hedeker and Gibbons, 1996). Difficulties can be encountered when the number of random effect parameters is not very small. With a detailed Monte Carlo experimentation, we show in this paper that the simulation-based estimators are almost as efficient as maximum likelihood. They are Simulated Maximum Likelihood (Gouriéroux and Monfort, 1991), Indirect Inference (Gouriéroux, Monfort and Renault, 1993) using an auxiliary approximated likelihood estimator, and Indirect Inference using an auxiliary linear probability model. The advantage of the latter is its great simplicity and computational speed.

# 1 Introduction

Random effects models are now widely used in applied Statistics and Econometrics. Applications include the analysis of the educational system (Aitkin and Longford, 1986) and the labour market (Mealli and Pudney, 1996; Pudney and Shields 2000). For the estimation of such models with binary responses various approximated solutions have been proposed, which make use of first or second-order Taylor expansion of the model (like the Marginal Quasi Likelihood estimator, MQL, proposed by Goldstein, 1991; see also Breslow and Clayton, 1993; Goldstein and Rasbash, 1996) or Approximated Likelihood proposed by Longford (1988). Numerical integration with Gaussian quadrature (Anderson and Aitkin, 1985; Hedeker and Gibbons, 1994) provides "in principle" an efficient solution to the estimation problem, but becomes very cumbersome with multiple level structures and covariates. Also the use of the EM algorithm (Stiratelli at al., 1984), the use of Gibbs sampling (Zeger and Karim, 1991), and Laplace approximations (Raudembush and Yang, 1998) have been proposed. Recently, Lee and Nelder (1996) introduced hierarchical or $h$-likelihood to estimate hierarchical linear models. The dispersion parameters are estimated by means of an adjusted profile $h$-likelihood, avoiding the integration that is necessary when marginal likelihood is used.

Simulation studies of Rodriguez and Goldman (1995) have shown the occurrence of large biases when estimation methods based on approximated likelihood are applied, especially in the binary case; recent works propose second order corrections (Goldstein and Rasbash, 1996) and bootstrap bias correction (Goldstein, 1996) to improve the estimates. Breslow and Lin (1995) and Lin and Breslow (1996) developed correction procedures for regression coefficients estimated by Penalized Quasi-Likelihod (PQL).

In this paper we propose three different simulation based estimators, and we compare their performances with the Maximum Likelihood estimator based on numerical integration (Anderson and Aitkin, 1985; Hedeker e Gibbons, 1994, 1996). The three methods are:

- Simulated Maximum Likelihood (Gouriéroux and Monfort, 1991);

- Indirect Inference (Gouriéroux et al. 1993; Gallant and Tauchen, 1996), which uses simulations performed under the initial model to correct the estimates derived from the approximated likelihood (MQL) auxiliary model (Longford, 1988; Mealli and Rampichini, 1999);
- Indirect Inference with a simple linear probability model as an auxiliary model.

Simulation-based methods are widely used in applied research. There is nevertheless a lack of studies on the relative efficiency of the alternative estimators.

With a detailed Monte Carlo experimentation, we show in this paper that the simulation-based estimators are almost as efficient as maximum likelihood.

Surprisingly, the indirect inference estimator based on a simple linear probability model has an excellent performance, with the advantage of great simplicity and computational speed.

Section 2 briefly illustrates the multilevel logit model. Section 3 describes alternative estimators for the model presented in section 2. Particularly, section 3.1 introduces the Maximum Likelihood estimator based on numerical integration, section 3.2 briefly describes the Simulated Maximum Likelihood estimator, section 3.3 presents the Indirect Inference estimator, and in particular section 3.3.2 proposes a procedure based on an approximated likelihood auxiliary estimator, while section 3.3.3 shows how a simple linear probability model can be used as an auxiliary model in Indirect Inference. In section 4 a Monte Carlo simulation study shows the performances of the various estimators. Section 5 concludes.

# 2 Multilevel Logit Models

We consider a two level logistic model. Given a $N \times 1$ vector of binary responses $\mathbf{y}$, with $y_{ik}$ elements, $i = 1, \cdots, n_k$; $k = 1, \cdots, K$; $N = \sum_{k=1}^{K} n_k$, we assume that, conditional on a vector of random effects $\mathbf{u}$, the elements of $\mathbf{y}$ are independent Bernoulli random variables with probabilities

$$\begin{aligned} p_{ik} &= Pr\{Y_{ik} = 1\} = \frac{\exp(\mathbf{x}_{ik}^{'}\boldsymbol{\gamma} + \mathbf{z}_{ik}^{'}\mathbf{u}_k)}{1 + \exp(\mathbf{x}_{ik}^{'}\boldsymbol{\gamma} + \mathbf{z}_{ik}^{'}\mathbf{u}_k)} \\ \text{logit}(p_{ik}) &= \ln\left(\frac{p_{ik}}{1 - p_{ik}}\right) = \mathbf{x}_{ik}^{'}\boldsymbol{\gamma} + \mathbf{z}_{ik}^{'}\mathbf{u}_k \end{aligned} \tag{2.1}$$

where $\mathbf{x}_{ik}$ is a $q$-vector of covariates having fixed effects $\boldsymbol{\gamma}$, $\mathbf{z}_{ik}$ is a $m$-vector of covariates having random effects $\mathbf{u}_k$, and the $\mathbf{u}_k \sim$ IID-$MN(\boldsymbol{\mu}, \Omega)$ are conditionally independent with the $\mathbf{x}$ variables, and $\Omega$ is an $m \times m$ covariance matrix . The binary response variable can be represented as the sum of the (2.1) probability and a residual $\varepsilon_{ik}$: $y_{ik} = p_{ik} + \varepsilon_{ik}$. Note that the residual can assume only the values $p_{ik}$ and $1 - p_{ik}$, it has zero mean and, given the value of the probability $p_{ik}$, his variance is equal to $p_{ik}(1 - p_{ik})$.

This formulation leads to the following logistic model for the $\mathbf{y}$ vector of observations

$$\mathbf{y} = h(\mathbf{X}\boldsymbol{\gamma} + \mathbf{Z}\mathbf{u}) + \varepsilon \tag{2.2}$$

where $h(.) = \text{logit}^{-1}(.)$ is the inverse of the link function (Mc Cullagh and Nelder, 1989), $\varepsilon$ is an error term with mean $\mathbf{0}$ and variance depending on $\mathrm{E}(\mathbf{y})$, $\mathbf{X}$ is the model matrix for the fixed effect $\boldsymbol{\gamma}$ and $\mathbf{Z}$ is the model matrix for the random effects $\mathbf{u}$.

Letting $\mathbf{y}_k$ denote the vector of responses for the $n_k$ units nested within the $k$-th group, the conditional density function has the binomial form

$$L(\mathbf{y}_k \mid \boldsymbol{\gamma}, \mathbf{u}_k) = \prod_{i=1}^{n_k} p_{ik}^{y_{ik}} (1 - p_{ik})^{1-y_{ik}} \tag{2.3}$$

while the conditional likelihood for all the groups is

$$L(\boldsymbol{\gamma} \mid \mathbf{u}) = \prod_{k=1}^{K} L(\mathbf{y}_k \mid \boldsymbol{\gamma}, \mathbf{u}_k) \tag{2.4}$$

To obtain the unconditional likelihood we need to integrate out the random effects $\mathbf{u}$.

The marginal density of $\mathbf{y}_k$ is expressed as the following integral of the density (2.3):

$$L(\mathbf{y}_k) = \int \cdots \int L(\mathbf{y}_k \mid \boldsymbol{\gamma}, \mathbf{u}_k) \phi(\mathbf{u}_k) \mathrm{d}u_1 \cdots \mathrm{d}u_m \tag{2.5}$$

where $\phi(.)$ denotes the multivariate normal density. Considering that observation of different groups are independent, the unconditional likelihood for all the groups can be written as the product of the $K$ independent marginal densities (2.5):

$$L(\boldsymbol{\gamma}, \Omega) = \prod_{k=1}^{K} L(\mathbf{y}_k) \tag{2.6}$$

and so the marginal log-likelihood for the $K$ groups becomes:

$$\log L(\boldsymbol{\gamma}, \Omega) = \sum_{k=1}^{K} \log L(\mathbf{y}_k) \tag{2.7}$$

# 3 Estimators

The intractable likelihood function (2.6) can be estimated via numerical integration (Hedeker and Gibbons, 1994, 1996). Numerical Maximum Likelihood (NML) gives consistent and efficient estimates of the parameters, but it becomes very cumbersome with multiple level structures and covariates. In the recent literature many approximated solutions of the intractable likelihood function (2.6) were proposed, which make use of first or second-order Taylor expansion of the model (2.2) (Goldstein, 1991, Breslow and Clayton, 1993; Goldstein and Rasbash, 1996), while Longford (1988, 1994) has proposed an approximation to the likelihood function (2.6).

In this paper we propose three simulation-based estimators: the Simulated Maximum Likelihood (SML) and two Indirect Inference (II) estimators. SML is, like NML, efficient, while the II estimators are not. Anyway, II is simpler and computationally less expensive

than SML, and seems to be almost efficient as the other, as our Monte Carlo simulation study shows.

Particularly, one of the two proposed Indirect Inference estimators is based on a full-iteration MQL estimator (II-MQL), while the other is based on a simple linear probability model as auxiliary model (II-OLS). We show that the II-OLS estimator performs as good as the more sophisticated II-MQL estimator (Mealli and Rampichini, 1999).

## 3.1 Maximum Likelihood via numerical integration

The Numerical Maximum Likelihood (NML) estimates of (2.5) can be obtained by means of a Fisher Scoring solution (Hedeker and Gibbons, 1996). For the scoring solution, the Cholesky factor of the random effects variance-covariance matrix is estimated, along with the coefficients of model covariates. In the estimation of the parameters for model (2.2), the authors suggest to orthogonally transform the response model. Let $\mathbf{u} = \mathbf{T}\boldsymbol{\alpha} + \boldsymbol{\mu}$, where $\mathbf{T}\mathbf{T}' = \Omega$ is the Choleski decomposition of $\Omega$. The reparametrization of model (2.2) is then

$$\mathbf{y} = h(\mathbf{X}\boldsymbol{\gamma} + \mathbf{Z}(\mathbf{T}\boldsymbol{\alpha} + \boldsymbol{\mu})) + \varepsilon$$

and the marginal density (2.5) becomes

$$L(\mathbf{y}_k) = \int \cdots \int L(\mathbf{y}_k \mid \boldsymbol{\gamma}, \boldsymbol{\alpha}_k)\phi(\boldsymbol{\alpha}_k)\mathrm{d}\alpha_1 \cdots \mathrm{d}\alpha_m \tag{3.1}$$

In order to solve the integrals that appear in the likelihood equations, Hedeker and Gibbons (1994) use numerical integration on the transformed $\boldsymbol{\alpha}$ space. For this, Gauss-Hermite quadrature can be used to approximate the integrals to any practical degree of accuracy. In Gaussian quadrature, the integration is approximated by a summation on a specified number of quadrature points $Q$ for each dimension of integration; thus, for the transformed $\boldsymbol{\alpha}$ space, the summation goes over $Q^r$ points, where $r$ is the number of covariates having random effects.

The MIXOR program (Hedeker and Gibbons, 1996) performs this kind of estimation for binary and ordinal response variables.

## 3.2 Simulated Maximum Likelihood

SML is a simulated version of ML. Given the ML estimator of a parameter $\theta$:

$$\hat{\theta} = \arg\max_{\theta} \log f(\mathbf{y} \mid \mathbf{X}; \theta) \tag{3.2}$$

where $f(\mathbf{y} \mid \mathbf{X}; \theta)$ is the probability density function (p.d.f.) of $\mathbf{y}$ given $\mathbf{X}$, suppose that, like in our case, this p.d.f. has an intractable form, while each conditional p.d.f. $f(\mathbf{y}_k, \mathbf{X}_k, \mathbf{u}; \theta)$, (see eq. 2.3) appearing in the decomposition:

$$\log f(\mathbf{y} \mid \mathbf{X}; \theta) = \sum_{k=1}^{K} \log f(\mathbf{y}_k \mid \mathbf{X}_k; \theta)$$

has a tractable form, and that the conditional distribution of $\mathbf{u}$ given $\mathbf{X}_k$ is known. Then we may draw independently, for each group $k$, $S$ simulated values $\mathbf{u}_k^s$, $s = 1, \cdots, S$ of the random terms $\mathbf{u}$. A simulated maximum likelihood estimator of $\theta$ is then obtained by replacing the unconditional p.d.f. with an unbiased approximation that is an empirical mean based on the simulated values:

$$\hat{\theta}_{SK} = \arg\max_{\theta} \sum_{k=1}^{K} \log \left[ \frac{1}{S} \sum_{s=1}^{S} f(\mathbf{y}_k, \mathbf{X}_k, \mathbf{u}_k^s; \theta) \right] \tag{3.3}$$

The SML estimator is consistent if $S$ and $K$ go to infinity, while the estimator is inconsistent if $S$ is fixed and $K$ goes to infinity. In practice, it is sufficient to retain a number $S$ of replications such that $\hat{\theta}_{SK} \simeq \lim_{S\to\infty} \hat{\theta}_{SK}$, so that $\hat{\theta}_{SK}$ is "close enough" to its limit value, and such a number often is of moderate size (Gourieroux and Monfort, 1996).

Moreover, if $S, K \to \infty$ and $\sqrt{K}/S \to 0$, then the SML estimator is asymptotically equivalent to the ML estimator (Gouriéroux and Monfort, 1991). Other estimators, like the simulated scores (Stern, 2000) do not require $S \to \infty$ for consistency. Performances of such an estimator are not examined in this paper and are left to future investigation.

## 3.3 The Indirect Inference procedure

Indirect Inference is an appropriate and natural procedure when direct inference is intractable, provided that the model of interest can be easily simulated for any fixed parameter value in the parameter space. According to Indirect Inference, estimation is made of two iterated steps. In the first one, an approximated (auxiliary) model or an auxiliary estimator is used to derive estimates of some auxiliary parameter. In the second step, simulations of the model of interest are used to correct the discrepancy of the auxiliary parameters from the original ones. We propose an Indirect Inference procedure using as the auxiliary estimator alternatively the MQL Approximated Maximum Likelihood Estimator (Longford, 1994) or the OLS estimator derived from a simple linear probability model.

### 3.3.1 The method

Suppose the model of interest be represented as

$$y = f(\theta, e) \tag{3.4}$$

and assume that values of $y$ (say $\tilde{y}$) can be easily produced by simulation of (3.4) for any value of $\theta \epsilon \Theta$, after drawing pseudo-random values of $e$ (say $\tilde{e}$). This model cannot be easily estimated, so we replace it with an approximation like

$$y = q(\beta, \eta) \tag{3.5}$$

which can be easily estimated.

For any $\theta \epsilon \Theta$, estimation of (3.5) with values $\tilde{y}$ generated from (3.4) leads to an estimator of $\beta$, $\beta(\theta)$, which depends on $\theta$ (and of course depends on the random errors, so it would be more appropriate $\beta(\theta, \tilde{e})$, but we omit $\tilde{e}$ from notation for simplicity).

The basic idea is to run $H$ simulations of the model of interest, calibrating $\theta$ until the estimates of the auxiliary parameters using simulated data are close, in some metric, to those using the actual data at hand.

For identification reasons, the dimension of the auxiliary parameter-vector must be at least as large as that of the parameters of interest. Gourieroux and Monfort (1996) give some intuitive guidelines on how to choose auxiliary models. The idea is to calibrate the $\theta$ parameters, keeping the drawings $\tilde{e}$ fixed, till we find the value $\hat{\theta}$ so that

$$\hat{\theta}(\Sigma_1) = \arg\min_{\theta} \left[\hat{\beta} - \beta(\theta)\right]' \Sigma_1 \left[\hat{\beta} - \beta(\theta)\right] \tag{3.6}$$

where $\hat{\beta}$ is the estimate of $\beta$ using the actual data, and $\Sigma_1$ is a symmetric positive-definite matrix, defining the metric. An alternative procedure was proposed by Gallant and Tauchen (1996)

$$\hat{\theta}(\Sigma_2) = \arg\min_{\theta} \left[\frac{\partial \log L(\tilde{y}, \beta)}{\partial \beta'}\right]_{\hat{\beta}} \Sigma_2 \left[\frac{\partial \log L(\tilde{y}, \beta)}{\partial \beta}\right]_{\hat{\beta}} \tag{3.7}$$

where $L$ is the pseudo-likelihood of the auxiliary model, and $\tilde{y}$ are values of $y$ produced by simulation of (3.4). In our case there is a one-to-one correspondence between the $\theta$ and $\beta$ parameters (exact identification), so that the estimates $\hat{\theta}(\Sigma_1)$ or $\hat{\theta}(\Sigma_2)$ do not depend on $\Sigma_1$ and $\Sigma_2$.

In a just-identified context, minimizing (3.6) is equivalent to finding $\hat{\theta}$ so that

$$\beta(\hat{\theta}) = \hat{\beta}. \tag{3.8}$$

At each step $\beta(\hat{\theta})$ is obtained as the mean of $H$ estimates of $\beta$ using $H$ independent simulated data of length $N$.

As equations (3.8) are only implicitly defined, in the case of exact identification (i.e. $\dim(\theta)$=$\dim(\beta)$) one can solve them using a search mechanism that defines a new trial of $\theta$ as

$$\theta^{j+1} = \theta^j - \beta(\theta^j) + \hat{\beta}. \tag{3.9}$$

More efficient algorithms can be used, such as

$$\theta^{j+1} = \theta^j + A\left[\beta(\theta^j) - \hat{\beta}\right] \tag{3.10}$$

with a careful choice of matrix $A$, which should determine the direction and the step size of the $j$-th step (An and Liu, 1996). In our application we take $A$ as the identity matrix; improvements, still under study, involve taking $A$ as the Jacobian matrix of derivatives of the auxiliary parameters w.r.t. the parameters of interest (see Calzolari, Di Iorio and Fiorentini, 1998). In the last case the iteration procedure described by the (3.9)-(3.10), is the well-known Newton-Raphson algorithm for solving a system of non-linear equations.

As a starting point for the procedure we use $\theta^0 = \hat{\beta}$. This is an obvious choice in our case, as there is a one-to-one correspondence between the $\theta$ and $\beta$ parameters. Perturbations of the initial values do not seem to affect the convergence of the procedure.

Minimizing (3.7) as from the method by Gallant and Tauchen (1996) in our just-identified case is equivalent to finding $\hat{\theta}$ such that $[\partial \log L(\tilde{y}, \beta)/\partial\beta]_{\hat{\beta}} = 0$, where $\tilde{y}$ are produced by simulation of (3.4) with $\hat{\theta}$ and fixed $\tilde{e}$. This gives *numerically* the same result as minimizing (3.6), but the computational procedure could be even faster, as it simply needs the computation of the gradient instead of the whole maximization with the simulated data. However, in this case the $A$ matrix cannot be simply the identity matrix, and the Jacobian we need to use contains derivatives of the gradient with respect to the $\theta$ parameters ($\partial[\partial \log L/\partial\beta]/\partial\theta'$).

As *auxiliary* estimators of model (2.2) we used the approximated maximum likelihood (or Marginal Quasi Likelihood, MQL) by Longford (1994) presented in the following section 3.3.2, and the least squares estimator of the linear probability model of section 3.3.3. In both cases, there is a straightforward one-to-one correspondence between the $\theta$ parameters ($\boldsymbol{\gamma}$ fixed coefficients and elements of $\Omega$ in the model of interest) and the $\beta$ parameters (coefficients and elements of the auxiliary covariance matrix $\Sigma$).

### 3.3.2 Maximizing approximated likelihood as an *auxiliary* estimator

An approximated solution of (2.6) can be obtained considering a second-order Taylor series expansion of the logarithm of the conditional likelihood (2.4) about $\mathbf{u} = \mathbf{0}$; the following approximate log-likelihood is obtained (Longford, 1994)

$$\log L(\boldsymbol{\gamma}, \Omega) \approx \log L(\boldsymbol{\gamma} \mid \mathbf{0}) - \frac{1}{2}\log|\mathbf{G}| + \frac{1}{2}\mathbf{e}'(\mathbf{W}_0 - \mathbf{V}_0^{-1})\mathbf{e} \tag{3.11}$$

where $\mathbf{G} = \mathbf{I} + \mathbf{Z}'\mathbf{W}_0\mathbf{Z}\Omega$, $\mathbf{e} = \mathbf{W}_0^{-1}(\mathbf{y} - \mathbf{h}_0)$, $\mathbf{h}_0 = \text{logit}^{-1}(\mathbf{X}\boldsymbol{\gamma})$, that is the inverse link evaluated at $\mathbf{u} = \mathbf{0}$, $\mathbf{W}_0$ is a diagonal matrix with entries $\mathbf{h}_0(1 - \mathbf{h}_0)$, $\mathbf{V}_0 = \mathbf{Z}\Omega\mathbf{Z}' + \mathbf{W}_0^{-1}$. Longford proposes to apply a Fisher scoring algorithm using the following derivatives:

1. derivatives of the log-likelihood (3.11) with respect to $\boldsymbol{\gamma}$, ignoring the dependence of $\mathbf{W}_0$ on $\boldsymbol{\gamma}$

$$\begin{aligned} \frac{\partial(\log L)}{\partial\boldsymbol{\gamma}} &\approx \mathbf{X}'\mathbf{V}_0^{-1}\mathbf{e} \\ -\frac{\partial^2(\log L)}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'} &\approx \mathbf{X}'\mathbf{V}_0^{-1}\mathbf{X} \end{aligned} \tag{3.12}$$

2. derivatives of the log-likelihood (3.11) with respect to a parameter $\omega\epsilon\Omega$

$$\begin{aligned} \frac{\partial(\log L)}{\partial\omega} &= \frac{1}{2}\left\{\mathbf{e}'\mathbf{V}_0^{-1}\frac{\partial\mathbf{V}_0}{\partial\omega}\mathbf{V}_0^{-1}\mathbf{e} - \text{tr}\left(\mathbf{V}_0^{-1}\frac{\partial\mathbf{V}_0}{\partial\omega}\right)\right\} \\ -\text{E}\left(\frac{\partial^2\log L}{\partial\omega_i\partial\omega_j}\right) &= \frac{1}{2}\text{tr}\left(\mathbf{V}_0^{-1}\frac{\partial\mathbf{V}_0}{\partial\omega_i}\mathbf{V}_0^{-1}\frac{\partial\mathbf{V}_0}{\partial\omega_j}\right) \end{aligned} \tag{3.13}$$

3. Longford (1994) shows that the expected second mixed partial derivatives are 0.

Rodriguez and Goldman (1995) show that the estimates of fixed effects and variance components obtained from Goldstein and Longford approximations are biased whenever the random effects are sufficiently large or the number of observations within a given level of clustering is small. Breslow and Clayton (1993) found that first order Marginal Quasi Likelihood (MQL) proposed by Goldstein is equivalent to the approximations used by Longford, and that this procedure leads to estimates of fixed and random effects that are biased towards zero. Goldstein (1995) and Goldstein and Rasbash (1995) have thus developed improved linearising approximations (PQL), and show that, for model with an adequate number of level 1 units per level 2 unit, these give satisfactory results. Nevertheless, when the number of level 1 units per level 2 unit is small and for binary responses as in Rodriguez and Goldman (1995), there is still some underestimation. Goldstein (1996) proposes an iterative bootstrap procedure that yields asymptotically unbiased and consistent estimates for such models.

We propose Indirect Inference to correct for the asymptotic bias of the MQL Approximated Likelihood Estimator.

Using the MQL estimator as an auxiliary estimator for II leads to an asymptotic covariance matrix which can be consistently estimated as (Mealli and Rampichini, 1999):

$$\hat{W}(H) = \left(1 + \frac{1}{H}\right) \left[ \frac{\partial^2 \log L}{\partial\theta\partial\beta'} \left(\mathbf{y}^s(\hat{\theta}), \hat{\beta}\right) W^{-1} \frac{\partial^2 \log L}{\partial\theta'\partial\beta} \left(\mathbf{y}^s(\hat{\theta}), \hat{\beta}\right) \right]^{-1} \tag{3.14}$$

where

$$W = \frac{N}{S} \sum_{s=1}^{S} (W_s - \bar{W})(W_s - \bar{W})' \tag{3.15}$$

with

$$W_s = \frac{\partial \log L}{\partial \beta} [\mathbf{y}^s(\hat{\theta}), \hat{\beta}] \tag{3.16}$$

$$\bar{W} = \frac{1}{S} \sum_{s=1}^{S} W_s \tag{3.17}$$

where $\theta = \{\boldsymbol{\gamma}, \Omega\}$ of model (2.2) and $\beta = \{\boldsymbol{\gamma}, \Omega\}$ are the auxiliary parameters, $\hat{\theta}$ is the Indirect Inference estimate of $\theta$ and $\mathbf{y}^s(\hat{\theta})$ a sample of simulated values of $y$ of length $N$ based on the parameter $\theta$; $\frac{\partial^2 log L}{\partial\theta\partial\beta'}[\mathbf{y}^s(\hat{\theta}), \hat{\beta}]$ is the numerical derivative of (3.13) and (3.14) with respect to $\theta$ evaluated at $\hat{\theta}$, and $S$ is the number of replications, usually chosen equal to $H$.

### 3.3.3 OLS of a linear probability model as an auxiliary estimator

In this section, we use a linear probability auxiliary model:

$$y_{ik} = \mathbf{x}'_{ik}\delta + \mathbf{z}'_{ik}\mathbf{u}_k + w_{ik} \tag{3.18}$$

There is a straightforward one-to-one correspondence between the parameters of the original model (2.2) and the auxiliary model (3.18). An estimate of the coefficients ($\hat{\delta}$) can be obtained by ordinary least squares, while an estimate of the variance-covariance matrix (say $\hat{\Sigma}$) can be computed via a regression of within-group residual products on some functions of the $\mathbf{z}_{ik}$.

Two examples will better illustrate the procedure. Consider the following model with random intercept

$$y_{ik} = \delta_0 + \delta_1 x_{ik} + u_k + w_{ik}. \tag{3.19}$$

where $u_k \sim i.i.d. - N(0, \sigma^2)$ and $w_{ik} \sim i.i.d. - N(0, 1)$.

As for two observation $i$ and $i'$ from the same group $k$ the $cov(u_k + w_{ik}, u_k + w_{i'k}) = \sigma^2$, the mean of within-group covariances is an estimate of the variance of $u_k$, say $\hat{\sigma}^2$.

As a second example, consider the following model with random intercept and slope:

$$y_{ik} = \delta_0 + \delta_1 x_{ik} + u_{1k} x_{ik} + u_{2k} + w_{ik}. \tag{3.20}$$

where $(u_{1k}, u_{2k}) \sim i.i.d. - N_2(\mathbf{0}, \Sigma)$, and $\Sigma$ has elements $\sigma_{11}, \sigma_{12}, \sigma_{22}$, and $w_{ik}$ as above.

As $cov(u_{1k}x_{ik} + u_{2k} + w_{ik}, u_{1k}x_{i'k} + u_{2k} + w_{i'k}) = \sigma_{11}x_{ik}x_{i'k} + \sigma_{12}(x_{ik} + x_{i'k}) + \sigma_{22}$, estimates of $\sigma_{11}, \sigma_{12}, \sigma_{22}$ can be obtained with a OLS regression of the within-group product of the OLS residuals on the previous regressors.

The asymptotic covariance matrix of the II estimator proposed in this Section is estimated via the GMM Method (Greene, 2000, sec. 11.5).

Note that the well-known drawback of the linear probability model, that is that the expected value of $Y$ can be outside the range $[0, 1]$, is not relevant for our purposes. In fact, we use this auxiliary model "only" to obtain parameters estimates.

# 4 Simulation study

In order to illustrate the proposed procedures, we consider two specification (Model1 and Model 2) for the probabilities of the following (two-level) logit model:

$$\begin{aligned} y_{ik} &= \text{logit}^{-1}(p_{ik}) + \varepsilon_{ik} \\ .y_{ik} &\sim Binomial(1, p_{ik}) \\ \mathrm{E}(\varepsilon_{ik}) &= 0 \end{aligned} \tag{4.1}$$

where $y_{ik}$ is the $i$-th observation of the $k$-th group, $n_k$ is the number of observations of the $k$-th group, and the variance of the error terms $\varepsilon_{ik}$ depends on $\mathrm{E}(y_{ik})$, with $i = 1, 2, \cdots, n_k$, $k = 1, 2, \cdots; K$ and $N = \sum_{k=1}^{K} n_k$.

Model 1: variance-component model (random intercept)

$$\begin{aligned} \text{logit } (p_{ik}) &= \gamma_0 + \gamma_1 x_{ik} + u_k \\ u_k &\sim N(0, \omega^2) \end{aligned} \tag{4.2}$$

The vector of parameters of interest, $\theta$, thus contains $\gamma_0$, $\gamma_1$ and $\omega^2$.

Model 2: random slope

$$\begin{aligned} \text{logit } (p_{ik}) &= \gamma_0 + (\gamma_1 + u_{1k}) x_{ik} + u_{2k} \\ (u_{1k}, u_{2k}) &\sim i.i.d. - N_2(\mathbf{0}, \Omega) \end{aligned} \tag{4.3}$$

where $\Omega$ has elements $\omega_{11}$, $\omega_{12}$ and $\omega_{22}$, and $\theta$ contains $\gamma_0$, $\gamma_1$, $\omega_{11}$, $\omega_{12}$ and $\omega_{22}$.

Note that the models can alternatively be interpreted as logit models for panel data with individual random intercept (and slope). In that case $y_{ik}$ would be the $i-$th observation of the $k-$th individual, with $i = 1, ..., n$; $k = 1, ..., T$; $N = nT$.

As *auxiliary* estimators for the II procedure we used the approximated maximum likelihood (or MQL) by Longford (1994) of section 3.3.2 and the least squares estimator of the linear probability model of section 3.3.3. In both cases, there is a straightforward one-to-one correspondence between the $\theta$ parameters (fixed coefficients and elements of $\Omega$ in the model of interest) and the $\beta$ parameters (fixed coefficients and elements of the auxiliary covariance matrix $\Sigma$).

## 4.1 Simulated Maximum Likelihood algorithm

To apply the Simulated Maximum Likelihood procedure, we have implemented a Fortran code that does the following steps:

1. Generation of a sample of observations for the dependent variable $y_{ik}$ as in equation (4.1), on the basis of the model probabilities (4.2) or (4.3). This is the sample of *pseudo-observed* data.

2. Perform a first order MQL estimation, in order to assign a starting value to $\theta$, where $\theta$ include the fixed parameters $\boldsymbol{\gamma}$ and the elements of the covariance matrix of the random effects $\Omega$.

3. Generation of a sample of $SK$ pseudo-random variables (Bianchi *et al.*, 1978) $\tilde{e}_k \sim IID - N(0,1)$ or $(\tilde{e}_{1k}, \tilde{e}_{2k}) \sim IID - N(0, I_2)$, where $S$ is the number of simulations. The pseudo-random variables $\tilde{e}_k$ are transformed into the random effects $\tilde{u}_k$ via the estimated $\Omega$ matrix.

4. Computation of the conditional density (2.3).

5. Computation of the integrals in the marginal density (2.5) as average over the $S$ replications for each group:

$$\tilde{L}(\mathbf{y}_k) = \frac{1}{S} \sum_{s=1}^{S} L(\mathbf{y}_k \mid \gamma, \mathbf{u}_s) \tag{4.4}$$

6. Computation of the simulated unconditional log-likelihood:

$$\log \tilde{L}(\gamma, \Omega) = \sum_{k=1}^{K} \log \tilde{L}(\mathbf{y}_k) \tag{4.5}$$

7. Estimation of the parameters in (4.5) is obtained via the Newton-Raphson method. Note that steps 3-4 are repeated more than one time at each iteration of the Newton-Raphson algorithm, in order to obtain the numerical score vector and the numerical information matrix of the simulated likelihood (4.5).

   The implementation of the more efficient algorithm based on analytical derivatives of the simulated likelihood (4.5) is under study.

## 4.2 Indirect Inference algorithm

To apply the Indirect Inference procedure, we have implemented a Fortran code that does the following steps:

1. Generation of a sample of observations for the dependent variable $y_{ik}$ as in equation (4.1), on the basis of the model probabilities (4.2) or (4.3). This is the sample of *pseudo-observed* data.

2. Approximated likelihood estimates of the auxiliary parameters $\hat{\beta}$ are obtained via the Fisher scoring algorithm (maximization of (2.6)) in the case of the II-MQL estimator; for the OLS auxiliary estimator derived from the linear probability model we used the estimation strategy presented in section 3.3.3.

3. The starting value $\theta^0$ for the iterative calibration procedure is chosen equal to the $\hat{\beta}$ value.

4. Generation of a sample of $HK$ pseudo-random variables $\tilde{e}_k$ i.i.d. $N(0,1)$ or $(\tilde{e}_{1k}, \tilde{e}_{2k})$ i.i.d. $N(0, I_2)$ and $H$ samples of length $N$ $\tilde{v}_{ik}$ i.i.d. $Uniform(0,1)$ of pseudo-random terms, where $H$ is the number of simulations. The pseudo-random effects $\tilde{u}_k$ i.i.d. $N(0, \omega^2)$ or $(\tilde{u}_{1k}, \tilde{u}_{2k})$ i.i.d. $N(0, \Omega)$ are obtained transforming the pseudo-random variables $\tilde{e}$ via the estimated $\Omega$ matrix.

5. Computation of the probabilities $\tilde{p}_{ik}$ following the model (4.2) or (4.3) and generation of $H$ samples of length $N$ $\tilde{y}_{ik} \sim Binomial(1, \tilde{p}_{ik})$, with

$$\tilde{y}_{ik} = \begin{cases} 1 & \text{if} \quad \tilde{v}_{ik} < \tilde{p}_{ik} \\ 0 & \text{otherwise} \end{cases}$$

6. Estimation of the parameters $\beta(\theta)$ on the sample of $NH$ pseudo-random $\tilde{y}_{ik}$ via the Fisher scoring algorithm (maximization of (2.5)) or OLS (estimation of the linear probability model).

7. The two vectors of parameters $\hat{\beta}$ and $\beta(\theta)$ are compared. If the following condition holds for a given value of $\varepsilon$, the estimation procedure terminates,

$$(\hat{\beta} - \beta(\theta)) \leq \varepsilon \tag{4.6}$$

otherwise the values of the $\theta$ parameters are modified via the calibration algorithm shown in (3.9) and a new iteration of the procedure starts again from step 5. It is important to stress that the pseudo-random effects and the pseudo-random variates generated at step 4 remain fixed in all iterations. The values of the simulated samples $\tilde{y}_{ik}$ are a function of the changing values of the $\theta$ parameters.

## 4.3 Simulation Results

The number of Monte Carlo replications is 1000, for model 1 (Tables 1 and 3) as well as for model 2 (Tables 2 and 4). In each replication, the total number of observations is 5000. For model 1 the number of groups, $K$, is set to 500, and the number of observations per group is constant, $n_k$=10. For model 2 we have adopted 1000 groups ($K = 1000$), with $n_k = 5$.

Numerical Integration estimations are obtained via the MIXOR program (Hedeker and Gibbons, 1996).

In each table, the left hand column displays the *true* value of the parameters. The other columns display the estimated parameters by each estimator (the Monte Carlo average over 1000 replications and the Monte Carlo variances, in square brackets). Monte Carlo mean of the asymptotic variance estimates of the parameters is reported in round brackets (equation 3.14 for II-MQL, GMM estimation for the II-OLS, Inverse of the Hessian for SML).

The II estimates have been performed maximizing the Approximated Likelihood (equivalent to MQL, as noted in section 2) or by OLS applied to the linear probability model (Tables 2 and 4). The right hand columns of Tables 2 and 4 display the estimated parameters in the auxiliary model (the Monte Carlo average over 1000 replications and the Monte Carlo variances, in square brackets).

Several considerations clearly emerge from the results.

- The three simulation-based estimators give almost the same mean of the estimated parameters and present quite the same Monte Carlo variances of the estimates for the fixed parameters.

- Maximization of the approximated likelihood leads to biased (inconsistent) estimates, at least as far as the slope and the variance parameters are concerned; this is evident if we look at the average estimates of $\gamma_1$ and $\omega$'s in the last but one columns of Tables 2 and 4.

- With respect to MQL, Indirect Inference and SML reduce the bias (inconsistency). The price we pay is a considerable enlargement of the variance of the estimates,

Table 1: Model 1 – SML and NML estimators: mean estim. param. [Monte-Carlo var.] and (Monte Carlo mean of asymp. var.).

| *Par.* | *True* | SML S=100 | NML |
|---|---|---|---|
| $\gamma_0$ | 0 | $.22\times10^{-2}$ [$.25\times10^{-2}$] ($.20\times10^{-2}$) | $-.54\times10^{-2}$ [$.18\times10^{-2}$] ($.15\times10^{-2}$) |
| $\gamma_1$ | 1 | 0.996 [$.18\times10^{-2}$] ($.14\times10^{-2}$) | 0.987 [$.17\times10^{-2}$] ($.16\times10^{-2}$) |
| $\omega^2$ | 0.5 | 0.492 [$.51\times10^{-2}$] ($.38\times10^{-2}$) | 0.502 [$.30\times10^{-2}$] ($.28\times10^{-2}$) |
| N=5000; K=500 | | | Replications=1000 |

Table 2: Model 1 –II estimators: mean estim. param., [Monte-Carlo var.] and (Monte Carlo mean of asymp. var.). Auxiliary parameter estimates from approx. lik. and linear prob. model.

| *Par.* | *True* | II-MQL H=50 | II-OLS H=50 | *Par.* | MQL | *Par.* | OLS |
|---|---|---|---|---|---|---|---|
| $\gamma_0$ | 0 | $-.70\times10^{-3}$ [$.21\times10^{-2}$] ($.21\times10^{-2}$) | $-.49\times10^{-3}$ [$.21\times10^{-2}$] ($.21\times10^{-2}$) | $\gamma_0$ | $-.14\times10^{-3}$ [$.19\times10^{-2}$] | $\delta_0$ | .500 [$.82\times10^{-4}$] |
| $\gamma_1$ | 1 | 0.987 [$.16\times10^{-2}$] ($.16\times10^{-2}$) | 1.006 [$.16\times10^{-2}$] ($.19\times10^{-2}$) | $\gamma_1$ | .907 [$.12\times10^{-2}$] | $\delta_1$ | .191 [$.29\times10^{-4}$] |
| $\omega^2$ | 0.5 | 0.500 [$.57\times10^{-2}$] ($.53\times10^{-2}$) | .499 [$.57\times10^{-2}$] ($.62\times10^{-2}$) | $\omega^2$ | .327 [$.22\times10^{-2}$] | $\sigma^2$ | $.18\times10^{-1}$ [$.50\times10^{-5}$] |
| N=5000; K=500 | | | | | | | Replications=1000 |

particularly for the variance parameters. Note that although the the SML and II estimators are consistent, a small bias persists also for quite large samples, as we can note computing the 95% confidence interval for the point estimates presented in Tables 3 and 4.

- A small number of replications did not converge, and they have been discarded (less than 0.5% in each table).

- The OLS estimator of the linear probability model does not seem to be related to the parameters of interest. Surprisingly, however, it has excellent performances when it is used as auxiliary estimator in indirect inference. The estimates are no longer biased (inconsistent: and this is of course what we expected), but the variances are about the same as those of the indirect estimates based on MQL. And this result is obtained with a very strong simplification, and a computational time about 20 times shorter

Table 3: Model 2 – SML and NML estimators: mean estim. param. [Monte-Carlo var.] and (Monte Carlo mean of asymp. var.)

| *Par.* | *True* | SML S=100 | NML |
|---|---|---|---|
| $\gamma_0$ | 0 | $-.15\times10^{-1}$ $[.22\times10^{-2}]$ $(.17\times10^{-2})$ | $.44\times10^{-2}$ $[.16\times10^{-2}]$ $(.18\times10^{-2})$ |
| $\gamma_1$ | 1 | 0.999 $[.35\times10^{-2}]$ $(.30\times10^{-2})$ | 1.00 $[.26\times10^{-2}]$ $(.30\times10^{-2})$ |
| $\omega_{11}$ | 0.5 | 0.486 $[.48\times10^{-2}]$ $(.77\times10^{-2})$ | 0.505 $[.87\times10^{-2}]$ $(.11\times10^{-1})$ |
| $\omega_{12}$ | 0.3 | 0.305 $[.42\times10^{-2}]$ $(.51\times10^{-2})$ | 0.313 $[.50\times10^{-2}]$ $(.52\times10^{-2})$ |
| $\omega_{22}$ | 0.5 | 0.489 $[.11\times10^{-1}]$ $(.12\times10^{-1})$ | 0.509 $[.11\times10^{-1}]$ $(.13\times10^{-1})$ |
| N=5000; K=1000 | | | Replications=2000 |

Table 4: Model 2 – II estimators: mean estim. param., [Monte-Carlo var.] and (Monte Carlo mean of asymp. var.). Auxiliary parameter estimates from approx. lik. and linear prob. model.

| *Par.* | *True* | II-MQL H=50 | II-OLS H=50 | MQL *Par.* | | OLS *Par.* | |
|---|---|---|---|---|---|---|---|
| $\gamma_0$ | 0 | $-.24\times10^{-2}$ $[.18\times10^{-2}]$ $(.19\times10^{-2})$ | $-.60\times10^{-3}$ $[.18\times10^{-2}]$ $(.18\times10^{-2})$ | $\gamma_0$ | $-.48\times10^{-1}$ $[.12\times10^{-2}]$ | $\delta_0$ | .488 $[.58\times10^{-4}]$ |
| $\gamma_1$ | 1 | 0.998 $[.30\times10^{-2}]$ $(.35\times10^{-2})$ | 1.000 $[.33\times10^{-2}]$ $(.35\times10^{-2})$ | $\gamma_1$ | .788 $[.14\times10^{-2}]$ | $\delta_1$ | .172 $[.44\times10^{-4}]$ |
| $\omega_{11}$ | 0.5 | 0.501 $[.83\times10^{-2}]$ $(.88\times10^{-2})$ | .502 $[.82\times10^{-2}]$ $(.88\times10^{-2})$ | $\omega_{11}$ | .255 $[.21\times10^{-2}]$ | $\sigma_{11}$ | $.176\times10^{-1}$ $[.70\times10^{-5}]$ |
| $\omega_{12}$ | 0.3 | 0.301 $[.55\times10^{-2}]$ $(.69\times10^{-2})$ | .301 $[.62\times10^{-2}]$ $(.69\times10^{-2})$ | $\omega_{12}$ | .154 $[.12\times10^{-2}]$ | $\sigma_{12}$ | $.743\times10^{-2}$ $[.30\times10^{-5}]$ |
| $\omega_{22}$ | 0.5 | 0.501 $[.15\times10^{-1}]$ $(.19\times10^{-1})$ | .506 $[.18\times10^{-1}]$ $(.19\times10^{-1})$ | $\omega_{22}$ | .215 $[.32\times10^{-2}]$ | $\sigma_{22}$ | $.104\times10^{-1}$ $[.56\times10^{-5}]$ |
| N=5000; K=500 | | | | | | | Replications=1000 |

- As expected, the NML presents the lowest asymptotic variance for the random parameter. The simulation-based estimators are about as efficient as maximum likelihood for the fixed parameters.
- The comparison of the two simulation-based estimators with the NML estimator is promising and clearly indicates the usefulness of these techniques for more compli-

cated models, where NML is most cumbersome.

- Note that our choice of an unusually large value $S = 100$ in SML was done only to show the equivalence with NML, which is ensured only for $S \to \infty$.

### 4.4 Computational remarks

Several problems arise in iterating to convergence the indirect estimation procedure for the models of this paper. When calibrating parameters, it can happen that a "very small" change of a $\theta$ parameter causes a "non-small" change in $\beta$ parameters. This is due to the discontinuous nature of the data: a simulated value of $y_{ik}$ equal to zero can become equal to one after a small change of $\theta$.

This has two main consequences:

- We cannot compute reliable values of the first derivatives of the auxiliary parameters with respect to the parameters of interest. It is therefore very hard to get convergence of the calibration procedure using the Jacobian matrix, as mentioned in section 3.1. Also, it is very difficult to apply the method by Gallant and Tauchen (1996). Thus our results have been obtained with the simple calibration rule of equation (3.9), even if it is clearly computationally inefficient.

- When the number of simulated data is not very large (5000, in our computations, when $H = 1$) it is very hard to ensure convergence with a reasonably tight convergence criterion. We found that it is quite difficult to go beyond the second significant decimal digit, or even worse: the first digit is usually guaranteed, but not always the second digit. Thus, results of indirect inference with $H = 1$ were not very accurate, and are not displayed here. Things go slightly better if we use a value of $H = 10$, as it is used in many practical applications (e.g. Calzolari, Fiorentini and Sentana, 2001). Things go much better when we use large values of $H$. For example, the results related to $H = 50$ have been obtained with a convergence criterion that ensures three decimal digits in each replication. Note that the value $H = 50$ is larger than values usually adopted in practical applications. With $H = 1$ the variance should be about the double than for $H = 50$, and this was not always clear from our results. To some extent, only results related to $H = 50$ are reliable.

## 5 Concluding remarks

In the paper we consider three simulation-based approaches as a tool to obtain a consistent estimator for logit models with random effects. The methods appear to be promising; they could be applied to other typical problematic situations, as is the case of a small number

of observations per group, multiple level structures and covariates, usually leading to large biases.

A remarkable result of this paper is the good performance of the Indirect Inference estimation technique even when a very rough auxiliary estimator is adopted. Our Monte Carlo experiment has in fact evidenced the same efficiency either when we use an approximated maximum likelihood (a good, but computationally complex auxiliary estimator), or when we use a linear probability model estimated by ordinary least squares (very poor, but computationally quite convenient).

# Acknowledgements

This research was supported by the University and Research Ministry of Italy (MURST) research projects "Stochastic models and simulation methods for dependent data" and "Evaluating quality, effectiveness and efficiency in individual services". The authors are grateful to people who commented the earlier version of the paper, including Bruno Chiandotto, Fabrizio Cipollini, Laura Grassini and two anonymous referees.

# References

An Y, Liu M. 1996. Using Indirect Inference to Solve the Initial Conditions Problem, paper presented at the *Optimizations Days Conference* (Montreal, Quebec, May 13-15).

Aitkin M, Longford N. 1986. Statistical Modelling Issues in School Effectiveness Studies, *Journal of the Royal Statistical Society A* **149**: 1–43.

Anderson DA, Aitkin M. 1985. Variance component models with binary response: interviewer variability, *Journal of the Royal Statistical Society B* **47**: 203–210.

Bianchi C, Calzolari G, Corsi P. 1978. A program for stochastic simulation of econometric models, *Econometrica* **46**: 235-236.

Breslow NE, Clayton DG. 1993. Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association* **88**: 9-25.

Breslow NE, Lin X. 1995. Bias Correction in generalized linear mixed models with a single component of dispersion, *Biometrika* **82**: 81-91.

Calzolari G, Di Iorio F, Fiorentini G. 1998. Control Variates for Variance Reduction in Indirect Inference: Interest Rate Models in Continuous Time, *Econometrics Journal* **1**: C100-C112.

Calzolari G, Fiorentini G, Sentana E. 2001. Constrained Indirect Inference Estimation, *Discussion Paper* **384**, Financial Market Group, London School of Economics.

Gallant AR, Tauchen G. 1996. Which Moments to Match?, *Econometric Theory* **12**: 657-681.

Goldstein H. 1991. Nonlinear multilevel models, with an application to discrete response data, *Biometrika* **78**: 45-51.

Goldstein H. 1996. Consistent estimators for multilevel generalized linear models using an iterated bootstrap, *Working Paper, Multilevel Models Project* (University of London).

Goldstein H, Rasbasch J. 1996. Improved approximations for multilevel models with binary responses, *Working Paper, Multilevel Models Project* (University of London).

Gori E, Mealli F, Rampichini, C. 1993. Indicatori di Efficienza ed Efficacia per la Valutazione dell'Attività di Formazione professionale, *Statistica* **3**: 501-533.

Gouriéroux C, Monfort A. 1991. Simulation Based Econometrics in Models with Heterogeneity, *Annales d'Economie et de Statistique* **20**: 69-107.

Gouriéroux C, Monfort A. 1996. *Simulation-based Econometric Methods*, Oxford University Press.

Gouriéroux C, Monfort A, Renault E. 1993. Indirect Inference, *Journal of Applied Econometrics* **8**: S85-S118.

Greene WH. 2000. *Econometric Analysis*, Fourth edition, Upper Saddle River, NJ: Prentice-Hall Inc.

Hedeker D, Gibbons RD. 1994. A random effects ordinal regression model for multilevel analysis, *Biometrics* **50**: 933-944.

Hedeker D, Gibbons RD. 1996. MIXOR A Computer Program for Mixed-effects ordinal probit and logistic regression analysis, *Computer Methods and Programs in Biomedicine* **49**: 157-176.

Lee Y, Nelder JA. 1996. Hierarchical Generalized Linear Models, *Journal of the Royal Statistical Society B*, **58**: 619-678.

Lin X, Breslow NE. 1996. Bias Correction in generalized linear mixed models with multiple components of dispersion, *Journal of the American Statistical Association* **91**: 1007-1016.

Longford NT. 1988. A quasi-likelihood adaptation for variance component analysis, *Proc. Statist. Comput. Sect. Am. Statist. Ass.*.

Longford NT. 1994. Logistic regression with random coefficients, *Journal of Computational Statistics & Data Analysis* **17**: 1-15.

Mealli F, Pudney SE. 1996. Occupational pensions and job mobility in Britain: estimation of random-effects competing risks model, *Journal of Applied Econometrics* **11**: 293-320.

Mealli F, Rampichini C. 1999. Estimating binary multilevel models through indirect inference, *Journal of Computational Statistics & Data Analysis* **29**: 313-324.

Pudney SE, Shields M. 2000. Gender, race, pay and promotion in the British nursing profession: estimation of the generalized ordered probit model, *Journal of Applied Econometrics*, to appear.

Raudembush S, Yang M. 1988. Maximum Likelihood for Hierarchical Models via High-order Multivariate Laplace Approximation, *mimeo*, University of Michingan.

Regione Toscana, Giunta Regionale, Osservatorio Regionale del mercato del lavoro. 1997. Monitoraggio e valutazione ex-post dell'attività di formazione professionale, anno formativo, 1993/94. Volume 1, *Rapporto Regionale, Flash Lavoro, Quaderni* **46**.

Rodriguez G, Goldman N. 1995. An Assessment of Estimation Procedures for Multilevel Models with Binary Responses, *Journal of the Royal Statistical Society A* **158**: 73-89.

Stiratelli R, Laird N, Ware JH. 1984 Random-effects models for serial observations with binary response, *Biometrics* **40**: 961 - 971.

Stern S. 2000. Simulation-based inference in econometrics: motivation and methods, in *Simulation-based inference in econometrics: methods and applications*, Mariano R, Schuermann T, Melvyn JW Eds., Cambridge: Cambridge University Press.

Yang M, Rasbash J, Goldstein H. 1998. MLwiN macros for advanced multilevel modelling. Technical Report, *Institute of Education*, University of London.

Zeger SL, Karim RM. 1991. Generalized linear models with random effects: a Gibbs sampler approach, *Journal of the American Statistical Association*, **86**: 79-86.