Multilevel models with

multivariate mixed response types

James Carpenter
London School of Hygiene & Tropical Medicine
Email: james.carpenter@lshtm.ac.uk
www.missingdata.org.uk

Support from MRC, ESRC & German Research Foundation

June 23, 2008

Acknowledgements

Harvey Goldstein (Bristol)

Mike Kenward (LSHTM)

Kate Levin (Edinburgh)

Based on: Goldstein H, Carpenter JR, Kenward MG and Levin KA (2008) Multilevel models with multivariate mixed response types. *Statistical Modelling (in press)*. Download from www.missingdata.org.uk.

Free goodies on line

From http://www.cmm.bristol.ac.uk/index.shtml you can download:

- Software: free standing executable program with
  - ASCII and worksheet input and output
  - Graphical menu based input specification
  - Model equation display
  - Monitoring of MCMC chains
- A training manual containing:
  - Outline of methodology
  - Worked through examples

Outline

- Example: childhood and adult height
- Model
- Estimation via MCMC
- More general model
- Notes on estimation
- Application to Multiple Imputation (MI)
- Example: Scottish Childhood ...
- Comparison with other approaches to MI
- Discussion.

Example: childhood and adult height

To illustrate the approach, consider modelling childhood and adult heights.

We have a two level model

- Level 1 is the repeated measures of childhood height
- Level 2 is the adult height

Such a model could be used to predict adult height from childhood height measurements.

Model

Let the superscript (1) denote level 1 (childhood heights) and (2) level 2 (adult height).

Let $j = 1, ..., J$ denote people; $i = 1, ..., I_j$ childhood height measurements.

Model:

$$y_i^{(2)} = \gamma_0 + u_{0j}^{(2)}$$

$$y_{ij}^{(1)} = (\beta_0 + u_{0j}^{(1)}) + (\beta_1 + u_{1j}^{(1)})t_{ij} + \beta_2 t_{ij}^2 + \beta_3 t_{ij}^3 + e_{ij}$$

$$\begin{pmatrix} u_{0j}^{(2)} \\ u_{0j}^{(1)} \\ u_{1j}^{(1)} \end{pmatrix} \sim MVN(0, \Omega_u)$$

$$e_{ij} \sim N(0, \sigma_e^2)$$

Estimation via Markov Chain Monte Carlo (MCMC)

- We could fit this model by maximum likelihood; but we use MCMC, because this generalises to more complex models more readily.

Estimation via Markov Chain Monte Carlo (MCMC)

- We could fit this model by maximum likelihood; but we use MCMC, because this generalises to more complex models more readily.
- In Bayesian inference we put a prior on the parameters, then calculate the posterior distribution of the parameters given the data.

Estimation via Markov Chain Monte Carlo (MCMC)

- We could fit this model by maximum likelihood; but we use MCMC, because this generalises to more complex models more readily.
- In Bayesian inference we put a prior on the parameters, then calculate the posterior distribution of the parameters given the data.
- If the prior is uninformative, and the posterior is approximately multivariate normal, inference is similar to classical approaches.

Estimation via Markov Chain Monte Carlo (MCMC)

- We could fit this model by maximum likelihood; but we use MCMC, because this generalises to more complex models more readily.
- In Bayesian inference we put a prior on the parameters, then calculate the posterior distribution of the parameters given the data.
- If the prior is uninformative, and the posterior is approximately multivariate normal, inference is similar to classical approaches.
- MCMC is a way of sampling from the posterior *without formally calculating it*.

Estimation via Markov Chain Monte Carlo (MCMC)

- We could fit this model by maximum likelihood; but we use MCMC, because this generalises to more complex models more readily.
- In Bayesian inference we put a prior on the parameters, then calculate the posterior distribution of the parameters given the data.
- If the prior is uninformative, and the posterior is approximately multivariate normal, inference is similar to classical approaches.
- MCMC is a way of sampling from the posterior *without formally calculating it*.
- We start an iterative simulation process under certain rules which guarantee that — after some initial iterations (the ‘burn in’) — the simulated draws come from the true posterior.

Estimation via Markov Chain Monte Carlo (MCMC)

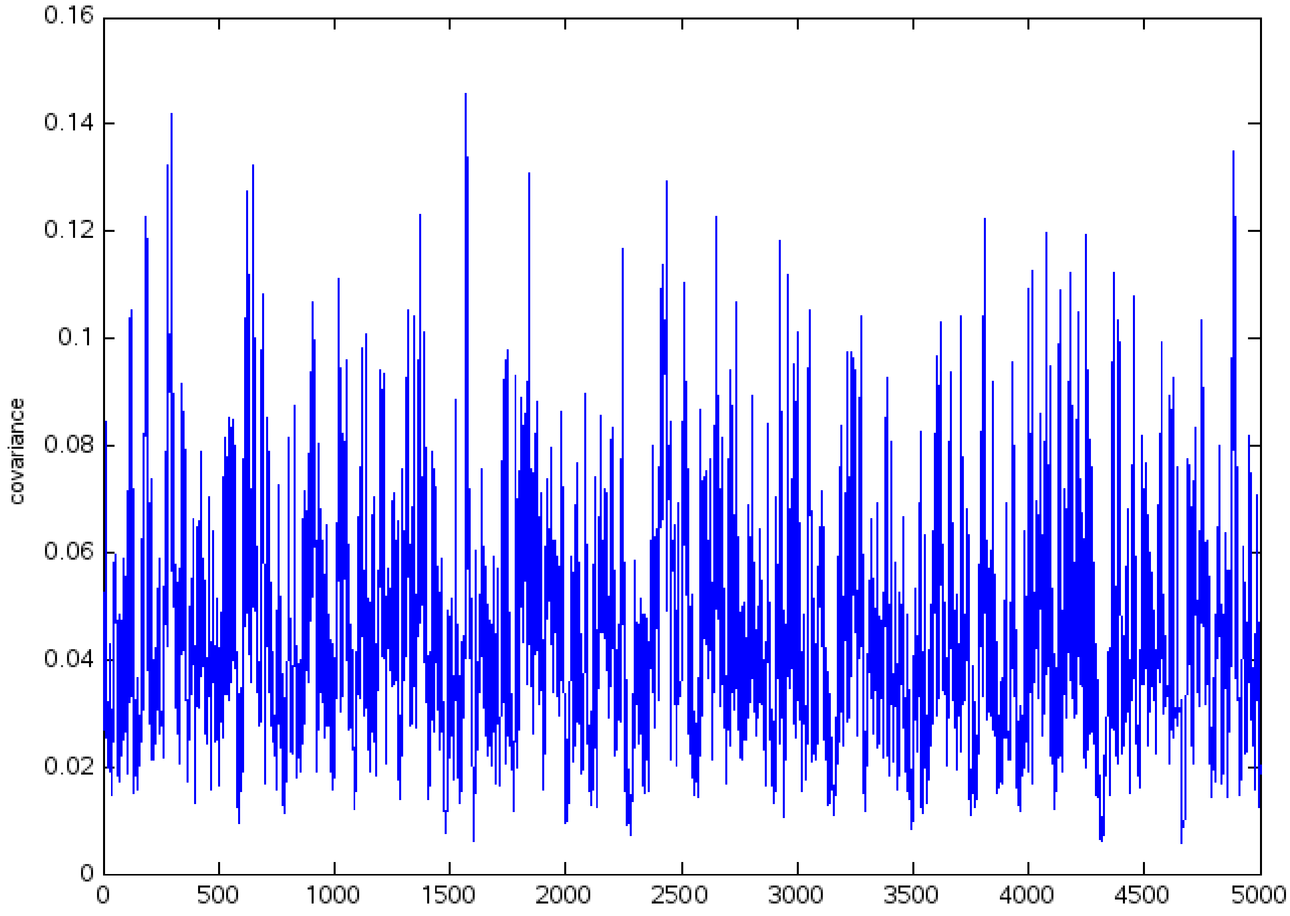
- We could fit this model by maximum likelihood; but we use MCMC, because this generalises to more complex models more readily.
- In Bayesian inference we put a prior on the parameters, then calculate the posterior distribution of the parameters given the data.
- If the prior is uninformative, and the posterior is approximately multivariate normal, inference is similar to classical approaches.
- MCMC is a way of sampling from the posterior *without formally calculating it*.
- We start an iterative simulation process under certain rules which guarantee that — after some initial iterations (the ‘burn in’) — the simulated draws come from the true posterior.
- We then use these to estimate the mean and variance of our parameters.

Estimation via Markov Chain Monte Carlo (MCMC)

- We could fit this model by maximum likelihood; but we use MCMC, because this generalises to more complex models more readily.
- In Bayesian inference we put a prior on the parameters, then calculate the posterior distribution of the parameters given the data.
- If the prior is uninformative, and the posterior is approximately multivariate normal, inference is similar to classical approaches.
- MCMC is a way of sampling from the posterior *without formally calculating it*.
- We start an iterative simulation process under certain rules which guarantee that — after some initial iterations (the 'burn in') — the simulated draws come from the true posterior.
- We then use these to estimate the mean and variance of our parameters.

Illustration

For example, the simulated draws for a parameter might look like this:



Parameter estimates

Coefficient	Estimate	Std. Err.	
Level 1 model			
Intercept	153.05	0.69	
Age (centred 13 years)	7.07	0.16	
Age-squared	0.294	0.054	
Age-cubed	−0.208	0.029	
Level 2 model			
Intercept	174.7	0.80	
Level 2 covariance matrix			
	57.77	1.30	50.01
	1.30	0.53	1.24
	50.01	1.24	69.42
Level 1 variance	3.21		

Including mixed responses

We now show how to extend this model to include binary, ordinal and unordered categorical data.

All these variables can be observed at either level 1 or level 2.

Besides modelling mixed response data, an important application of this model is *multiple imputation*, which we return to after we have described the model.

We first sketch our approach for binary and ordinal responses, and then describe how unordered categorical responses can be handled.

Binary data

For simplicity, just use the index j, and let z_j be a binary variable.

Define a latent variable y_j by

$$y_j > 0 \iff z_j = 1,$$

and write

$$y_j = \beta_0 + \beta_1 x_j + e_j, \quad e_j \sim N(0, \sigma_e^2)$$

Then

$$\begin{aligned}\Pr(z_j = 0) = \Pr\{e_j < -(\beta_0 + \beta_1 x_j)\} &= \int_{-\infty}^{-(\beta_0+\beta_1 x_j)} \phi(t)dt \\ &= \Phi\{-(\beta_0 + \beta_1 x_j)\}\end{aligned}$$

Using this formulation we can include binary data in the likelihood at the appropriate level.

Ordinal data

We can extend this approach to ordinal data. Suppose we have K categories.

Now let z_j be the ordinal variable, with $\Pr(z_j = k) = \pi_k$, $k = 1, \ldots, K$.

Let $\gamma_k = \sum_{k=1}^{K} \pi_k$ and relate γ_k to covariates through

$$\gamma_k = \int_{-\infty}^{\alpha_k - (\beta_0 - \beta_1 x_j)} \phi(t)dt, \quad k = 1, \ldots, K-1.$$

Using this formulation we can include ordinal data in the likelihood at the appropriate level.

Unordered categorical data

We use the maximum indicant model (Aitchison and Bennett, 1970). Assume a response is one of K categories, and let $z_{jk} = 1$ if individual j gives category k and 0 otherwise.

We only need to model $K-1$ categories. For each, we have a separate regression coefficient β_k relating covariates x_j to $\Pr(z_{jk} = 1)$. Following a similar approach to above let

$$y_{jk} = x_j\beta_k + e_{jk}, \quad k = 1, \ldots, K-1, \text{ where}$$

$$e_j \overset{iid}{\sim} N_{K-1}(0, I_{K-1}).$$

Then, for $k = 1, \ldots, (K-1)$,

$$\begin{aligned} \Pr(z_{jk} = 1) &= \Pr(y_{jk} > y_{jk'}, \text{ all } k' \neq k) \\ &= \Pr\{e_{jk} - e_{jk'} > x_j(\beta_{k'} - \beta_k)\}, \text{ all } k' \neq k \end{aligned}$$

and $\Pr(z_{jK} = 1) = \Pr(y_{jk'} < 0), k' = 1, \ldots, (K-1)$.

Using standard properties of the normal distribution, these can be calculated and the appropriate term included in the likelihood.

Estimation

We use an MCMC algorithm to fit this model.

This uses Gibbs sampling, where the parameters in the model (including the random effects) are divided up into groups.

We then sample from the conditional distribution of each parameter group (given current values of all the other parameters) in turn.

Some conditional distributions are known parametric distributions, so we can use their samplers.

Others are not, so we use a Metropolis-Hastings step.

Scottish Health Behaviour in School Children Study

1644 pupils in 75 primary schools filled in a survey relating to the heath behaviour. Each school also completed a questionnaire.

Response is frequency of fruit intake (6 ordinal categories).

Variables of interest: school involved in health promotion initiative; school involved in 'hungry for success' initiative; fruit available in school.

Possible confounders: sex, father's social class Carstairs index of social deprivation (for school).

Only Carstairs index complete; missingness in other variables from 1.2% to 13.6%.

Multilevel, mixed response data.

Missing data

What is MI?

Multiple imputation is a stochastic estimation technique for partially observed data sets.

It involves imputing ‘completed’ data sets, fitting the model to each imputed data set, and combining the results using certain rules.

Its attraction is that the rules are simple and general, so that once the imputation model is chosen the process is semi-automatic.

Application to MI

When analysing partially observed data, we need to think about the stochastic mechanism generating the missing data.

One important class is unintuitively called ‘Missing at Random’.

This says that, conditional on fully observed variables, the chance of seeing potentially missing values and the actual values are independent.

If we can assume MAR, then we can get valid inference from regression models where the partially observed variables are *responses*.

> We therefore fit our multilevel mixed response model to the observed data (treating all variables as responses) and impute the missing data.

Results (for variables of interest)

Estimates are log-odds-ratios for increased fruit intake, adjusting for father's SES

Variable	Obs data	MI — REALCOM
Girl	0.21 (0.06)	0.24 (0.05)
Health promotion	-0.59 (0.52)	-0.56 (0.50)
Hungry for success	0.14 (0.21)	0.20 (0.18)
Cannot buy fruit vs every day	0.14 (0.13)	0.08 (0.11)

Burn in & updates between imputations: 1000; 20 imputations

Slow mixing with the threshold parameters for the categorical data, but chain appears stationary.

Multilevel structure important educational data.

Comparison with other MI approaches

	Complexity			
Response type	Normal		Mixed response	
Data structure	Independent	Multilevel	Multilevel	Indep^t
Package				
Standalone	NORM	PAN	REALCOM	
SAS	NORM-port	—	—	IVE
STATA	NORM-port	—	—	ICE
R/S+	NORM-port	—	—	MICE
MLwiN	MCMC algorithm emulates PAN		+ 1–2 binary	

All methods: General missingness pattern

Relationships essentially normal/linear (except MLwiN, REALCOM)

Interactions must usually be handled by separate imputation

Shafer has package for general location model, but this has seen limited use

Chained equations has weaker theoretical basis, and does not readily extend to full multilevel structure.

Summary

- Building on similar models in the literature, we have developed a multilevel multivariate response model.
- We have described an MCMC algorithm for fitting this, and programmed it in the 2-level case.
- Further work is needed to improve the performance of the MCMC fitting algorithm.
- A key application is multiple imputation; we have illustrated its use with an analysis of multilevel mixed response data.
- Multilevel structure needs to be accounted for in imputation to avoid bias in parameter and variance estimates — and hence in imputation.
- Other applications, and extensions, are described in the paper (see slide 3 above for details of downloads available).