AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

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

*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.*

CANTER

DP 9814

Department of Economics

UNIVERSITY OF CANTERBURY

CHRISTCHURCH, NEW ZEALAND

ISSN 1171-0705

LIBRARY
WITHDRAWN
DEC 15 1998
GIANNINI FOUNDATION OF
AGRICULTURAL ECONOMICS

BAYESIAN ANALYSIS OF MULTIVARIATE COUNT DATA

Siddhartha Chib and Rainer Winkelmann

Discussion Paper

No. 9814

Department of Economics, University of Canterbury
Christchurch, New Zealand

Discussion Paper No. 9814

November 1998

BAYESIAN ANALYSIS OF MULTIVARIATE COUNT DATA

Siddhartha Chib and Rainer Winkelmann

Bayesian Analysis of Multivariate Count Data

SIDDHARTHA CHIB
Washington University, St. Louis, MO 63130, USA.

RAINER WINKELMANN
University of Canterbury, Christchurch, New Zealand.

September 1998

Abstract

This paper is concerned with the analysis of multivariate count data. A class of models is proposed, based on the work of Aitchison and Ho (1989), in which the correlation amongst the counts is represented by correlated, outcome-specific, latent effects. Several interesting special cases of the model are discussed and a tuned and efficient Markov chain Monte Carlo algorithm is developed to estimate the model. The ideas are illustrated with three real data examples of trivariate to sixteen variate correlated counts.

Keywords: Correlated count data; Markov Chain Monte Carlo; Metropolis-Hastings algorithm.

1 Introduction

Regression analysis of univariate count data has been the subject of a large and still growing literature (for recent book surveys, see Winkelmann, 1997, and Cameron and Trivedi, 1998) but the regression analysis of correlated counts involving multivariate measurements on a random cross-section of subjects or repeated measurements on a sample of subjects over time is less well developed. Many of the existing models for multivariate counts impose strong *a-priori* restrictions on the correlation structure between counts that are unlikely to hold in applications. Furthermore, most of the estimation methods are concentrated on the bivariate case.

Correlated count data arises in many situations and in many disciplines. Bivariate examples include the counts of surface and interiors faults in lenses (Aitchison and Ho, 1989), number of doctor consultations and the number of other ambulatory visits (Gurmu and Elder, 1998), the number of voluntary and involuntary job changes (Jung and Winkelmann, 1993) and the number of entries and exits to an industry (Mayer and Chappell, 1992). Examples of correlated counts arising from longitudinal measurements are discussed, for example, by Chib, Greenberg and Winkelmann (1998), Diggle, Liang and Zeger (1994) and Hausman, Hall and Griliches (1984).

In this paper we propose a class of models for multivariate counts, based on the work of Aitchison and Ho (1989), in which the correlation amongst the counts is represented by correlated, outcome-specific, latent effects. The presence of the latent effects implies that the likelihood function of the model has no closed form expression. We discuss estimation of the model by Markov chain Monte Carlo simulation methods and demonstrate its efficacy in problems of upto sixteen dimensions. As far as we are aware a general correlated count data model with these many correlated counts has never been fit in the literature. The mixing properties of the algorithm are excellent in the examples.

The paper is organized as follows. In Section 2 we present the model and present some special cases and extensions. The fitting algorithm is developed in Section 3 while Section 4 gives three real data examples. Section 5 concludes.

2 Model

For the ith subject, let $y_i = (y_{i1}, ...y_{iJ})'$ be a vector of responses on a set of J count variables. The model of interest specifies that conditionally on response specific coefficients $\beta = (\beta_1, ..., \beta_J)$ (where each β_j is a vector of coefficients) and *latent* correlated random variables $e_i = (e_{i1}, \ldots, e_{iJ})$ the jth count is distributed as Poisson with parameter θ_{ij}:

$$y_{ij}|\theta_{ij} \sim \text{Poisson}(\theta_{ij}), \tag{1}$$

where

$$\theta_{ij} = \text{E}(y_{ij}|\beta_j, e_{ij}) = \exp(x_{ij}'\beta_j + e_{ij}),$$

and

$$e_i \sim \mathcal{N}_J(-0.5\text{diag}D, D),$$

a multivariate normal distribution with mean vector $-0.5\text{diag}D$ and covariance matrix D. Note that the diagonal elements of D appear in the mean specification.

This model is closely related to that proposed by Aitchison and Ho (1989). If we let $v_{ij} = \exp(e_{ij})$ and let $v_i = (v_{i1}, ..., v_{iJ})$, then our assumption on e_i implies that $v_i \sim \text{LN}_J(\mathbf{1}, \Sigma)$, a multivariate log-normal distribution with parameters $\mathbf{1}$ (a vector of ones) and dispersion matrix Σ where $\sigma_{ij} = \exp(d_{ij}) - 1$ and thus $\Sigma = \exp(D) - \mathbf{1}\mathbf{1}'$. Under this parameterization, $y_{ij}|\lambda_{ij} \sim \text{Poisson}(v_{ij}\lambda_{ij})$, where $\lambda_{ij} = \exp(x'_{ij}\beta_j)$. This is precisely the model in Aitchison and Ho (1989) except that it has been generalized so as to allow for response specific covariates x_{ij} and response-specific parameters β_j.

Now let $\lambda_i = (\lambda_{i1} \cdots \lambda_{iJ})'$ and $\Lambda_i = \text{diag}(\lambda_i)$, where $\lambda_{ij} = \exp(x'_{ij}\beta_j)$. Then by the law of the iterated expectation we get that

$$\text{E}(y_i|\beta, D) = \lambda_i \tag{2}$$

and

$$\text{Var}(y_i|\beta, D) = \Lambda_i + \Lambda_i[\exp(D) - \mathbf{1}\mathbf{1}']\Lambda_i \tag{3}$$

Hence, the covariance between the counts is represented by the terms

$$\text{Cov}(y_{ij}, y_{ik}) = \lambda_{ij}(\exp(d_{jk}) - 1)\lambda_{ik}, \ j \neq k$$

which can be *positive* or *negative* depending on the sign of d_{jk}, the (j,k) element of D.

Special cases of the model

This model formulation encompasses a variety of interesting sub-cases that are relevant in practice and have routine solutions when the dependent variable is continuous and the model is linear in the parameters, but not when the dependent variable is a count. Our initial interpretation is that i denotes individuals, while j indexes different characteristics, all measured as counts, for the same individual. One example occurs when studying the provision of health services where a researcher might be interested in a joint analysis of the number of individual visits to a doctor and to a non-doctor health specialist (Gurmu and Elder, 1998).

Seemingly unrelated regression. The general formulation can be easily transformed into the seemingly unrelated regression (SURE) model where the researcher has access to a cross-section of time series and y_{ij} measures the same characteristic for all i and j. In this context, the first index i represents time while the second index j represents the cross-section unit. $\text{Var}(y_i)$ is now a contemporaneous variance-covariance matrix and the diagonal elements allow for heteroscedasticity and overdispersion. The covariates x_{ij} may or may not vary in the cross-section. While in the linear model, system estimation increases efficiency only when $x_{ij} \neq x_{ik}$ for some $j \neq k$, this requirement does not apply in the case of multivariate counts. Alternative interpretations of the SURE model are possible. For instance, Aitchison and Ho (1989) give data by Arbous and Kerrich on measurements from three air samplers at 50 different locations. Here, $\text{Var}(y_i)$ accounts for the correlations between the measurements of the three samplers at a given location.

Panel models. The proposed multivariate Poisson model can also be used when data form an independent cross-section of time series. In particular, assume that data are reorganized so that i denotes the cross-section unit and j denotes time. Now, $\text{Var}(y_i)$ captures serial correlation for observations of cross-section unit i over time whereas there is no correlation, contemporaneous or else, between cross-sections. A special case arises when $e_{ij} = u_i + v_{ij}$ where u_i and v_{ij} are independent error components with constant variance. This "one-factor" approach reduces the number of free parameters in D from $J(J+1)/2$ to 2 and, as in the linear error components model, $D = \sigma_v^2 I_J + \sigma_u^2 \mathbf{1}\mathbf{1}'$.

In addition, the level of generality of the model is affected by whether the parameters β are heterogeneous or homogeneous. For instance, consider the following three conditional expectation functions:

$$\lambda_{ij}^1 = \exp(x_{ij}'\beta)\,, \quad \lambda_{ij}^2 = \exp(x_{ij}'\beta_i)\,, \quad \lambda_{ij}^3 = \exp(x_{ij}'\beta_j)\,.$$

In the second case, the parameters are allowed to vary over individuals, whereas in the third case, the parameters are allowed to vary over time.

Independent observations. There are two possibilities. First, assume that D is a diagonal matrix. The joint density of the counts on subject i , $f(y_i|\beta, D)$, collapses into a product of one dimensional integrals

$$p(y_i|\beta, D) = \prod_{j=1}^{J} \int f(y_{ij}|\beta_j, e_{ij})\phi(e_{ij}| - 0.5d_{jj}, d_{jj})\mathrm{d}e_{ij} \tag{4}$$

where f is the Poisson mass function with mean $\theta_{ij} = \exp(x'_{ij}\beta_j + e_{ij})$. Effectively, this is the product of J independent univariate Poisson-log normal densities as discussed, for instance, in Hinde (1982). A particular feature of this model is that it allows for data with extra-Poisson variation, or overdispersion. Alternatively, assume that $D = 0$. Then the joint density for y_i simplifies to a product of J independent Poisson densities.

The univariate Poisson model. For $J = 1$ the general model collapses to a univariate Poisson model with unobserved heterogeneity. If furthermore $D = 0$, this is the standard log-linear Poisson regression model.

3 Bayesian inference

3.1 Prior distributions

We suppose that the parameters (β, D) independently follow the prior distributions

$$\beta \sim N(\beta_0, B_0^{-1}), \quad D^{-1} \sim \mathrm{Wish}(\nu_0, R_0),$$

with density $\pi(\beta)\pi(D^{-1})$, where (β_0, B_0, v_0, R_0) are known hyperparameters and Wish $(\cdot,\cdot)$ is the Wishart distribution with ν_0 degrees of freedom and scale matrix R_0.

3.2 Likelihood function

Under conditional independence across subjects, the likelihood function is the product of the contributions $p(y_i|\beta, D)$, where $p(y_i|\beta, D)$ is the joint probability of the J counts in cluster i and is given by

$$p(y_i|\beta, D) = \int \prod_{j=1}^{J} f(y_{ij}|\beta_j, e_{ij})\phi(e_i| - 0.5\mathrm{diag}D, D)\mathrm{d}e_i \tag{5}$$

where f as above is the Poisson mass function distribution conditioned on (β_j, e_{ij}) and ϕ is the J-variate normal distribution. This J-dimensional integral cannot be solved in closed form. We therefore turn to Markov chain Monte Carlo methods to simulate the augmented posterior distribution (which is proportional to the product of the prior $\pi(\beta)\pi(D^{-1})$ and the joint density of the observations and the latent variables $\prod_{i=1}^{n} p(y_i|\beta, D, e_i)p(e_i|D)$, where $p(e_i|D)$ is the Gaussian density $\phi(e_i| - 0.5\text{diag}D, D)$).

3.3 MCMC implementation

To develop an operational Markov chain Monte Carlo scheme we follow Chib, Greenberg and Winkelmann (1998) and block the parameters as e, β, and D and recursively sample the full conditional distributions

$$[e|y, \beta, D]; \quad [\beta|y, e]; \quad [D^{-1}|e], \tag{6}$$

using the most recent values of the conditioning variables at each step. The details of the simulations are discussed next.

3.4 Sampling e

The target density is $\pi(e|y, \beta, D) = \prod_{i=1}^{n} \pi(e_i|y_i, \beta, D)$ which factors into the product of n independent terms. To sample the ith target density

$$\begin{aligned} \pi(e_i|y_i, \beta, D) &= c_i\phi(e_i|D) \prod_{j=1}^{J} \exp[-\exp(x'_{ij}\beta_j + e_{ij})] \, [\exp(x'_{ij}\beta_j + e_{ij})]^{y_{ij}} \\ &\equiv c_i\pi^*(e_i|y_i, \beta, D) \end{aligned} \tag{7}$$

we can utilize the Metropolis-Hastings algorithm [see for example Hastings (1970), Chib and Greenberg (1995)]. The proposal density is taken to be multivariate-t with parameters that are tailored to those of the target $\pi(e_i|y_i, \beta, D)$. Let $\hat{e}_i = \arg\max \ln \pi^*(e_i|y_i, \beta, D)$ and $V_{e_i} = (-H_{e_i})^{-1}$ be the inverse of the Hessian of $\ln \pi^*(e_i|y_i, \beta, D)$ at the mode $\hat{e}_i$. These quantities are obtained from a few Newton-Raphson steps using the gradient and Hessian matrix

$$g_{e_i} = -D^{-1}(e_i + \text{diag}D) + [y_i - \exp(x_i\beta + e_i)] \tag{8}$$

and

$$H_{e_i} = -D^{-1} - \text{diag}\{\exp(x_i\beta + e_i)\} \tag{9}$$

Then, our proposal density is given by $q(e_i|y_i, \beta, D) = f_T(e_i|\hat{e}_i, V_{e_i}, \nu)$, a multivariate-$t$ density with ν degrees of freedom (where ν is a tuning parameter). We now draw a proposal value $e_i^\dagger$ from $q(e_i|y_i, \beta, D)$ and move to $e_i^\dagger$ from the current point e_i with probability

$$\alpha(e_i, e_i^\dagger|y_i, \beta, D) = \min\left\{\frac{\pi^*(e_i^\dagger|y_i, \beta, D)q(e_i|y_i, \beta, D)}{\pi^*(e_i^\dagger|y_i, \beta, D)q(e_i^\dagger|y_i, \beta, D)}, 1\right\}. \tag{10}$$

If the proposal value is rejected, then the next item in the chain is the current value e_i.

3.5 Sampling β and D

Given e, we next sample β from the target density which is proportional to

$$\pi^*(\beta|y, e, D) = \phi(\beta|\beta_0, B_0^{-1}) \prod_{i=1}^{n} \prod_{j=1}^{J} \exp[-\exp(x_{ij}'\beta_j + e_{ij})] [\exp(x_{ij}'\beta_j + e_{ij})]^{y_{ij}}. \tag{11}$$

This density is sampled by the M-H algorithm in a manner that is analogous to that of e_i. The proposal distribution is based on the mode $\hat{\beta}$ and curvature $V_\beta = [-H_\beta]^{-1}$ of $\log \pi^*(\beta|y, e, D)$ where these quantities are found as before using a few Newton-Raphson steps with the gradient vector $-B_0(\beta - \beta_0) + \sum_{i=1}^n \sum_{j=1}^J [y_{ij} - \exp(x_{ij}'\beta_j + e_{ij})]x_{ij}$ and Hessian matrix $H_\beta = -B_0 - \sum_{i=1}^n \sum_{j=1}^J [\exp(x_{ij}'\beta_j + e_{ij})]x_{ij}x_{ij}'$.

Following Chib, Greenberg and Winkelmann (1998), we obtain the proposal by reflecting the current value around the modal value $\hat{\beta}$ and then add a Gaussian increment with variance $\tau_\beta V_\beta$. The resulting proposal density is $q(\beta, \beta^\dagger|y, e, D) = \phi(\hat{\beta} - (\beta - \hat{\beta}), \tau_\beta V_\beta)$ and the probability of move is given in terms of the ratio of density ordinates

$$\alpha(\beta, \beta^\dagger|y, e, D) = \min\left\{\frac{\pi^*(\beta^\dagger|y, e, D)}{\pi^*(\beta|y, e, D)}, 1\right\}, \tag{12}$$

since the proposal density is symmetric in $(\beta, \beta^\dagger)$ and hence cancels.

Finally, the sampling of D^{-1} is from $\pi(D^{-1}|b) = f_W(D^{-1}|n + v_0, [R_0^{-1} + \sum_{i=1}^n (\bar{e}_i\bar{e}_i')]^{-1})$, a Wishart density with $n + v_0$ degrees of freedom and scale matrix $[R_0^{-1} + \sum_{i=1}^n (\bar{e}_i\bar{e}_i')]^{-1}$ where $\bar{e}_i = e_i - \text{diag}D$.

4 Examples

We illustrate the use of the proposed algorithm on three different data sets. In the first example, taken from Aitchison and Ho (1989), we are concerned with the trivariate distribution of the number of bacterial counts for three samplers in 50 different locations. Apart from a response-specific intercept this examples does not contain any other covariates. Our second example is based on a longitudinal dataset by Diggle, Liang and Zeger (1994) on seizure counts for 59 epileptics over five time periods. The estimated model includes three covariates and five period-specific latent effect. Finally, the third example is considered with a very high dimensional problem on the number of airline incidents of sixteen U.S. passenger air carriers between 1957 and 1986.

Our algorithm was run for 6000 iterations following a burn-in phase of 500 iterations. In each example the results were found to be robust to the starting values (which for β was taken to be the ML estimate from independent Poisson regressions and for D was .1 times the identity matrix) and the tuning constants in the Metropolis-Hastings steps. In effect, the algorithm was applied with no user-intervention beyond the specification of the model and prior hyperparameters.

4.1 Bacterial colony counts

The first example has the structure of a seemingly unrelated regression (SURE) model, where bacterial colony counts for three different air samplers measured at the same 50 locations are potentially not independent. Correlation between the counts from the three samplers at a particular location can arise from common location specific variations in bacterial infestation. In such a situation, joint estimation will increase efficiency.

The prior parameters are set to

$$\beta_0 = 0,\ B_0^{-1} = 0.01 I_3,\ \nu_0 = 6,\ R_0 = I_3$$

The two scale factors are equal to unity. The marginal posterior distributions of β_1, β_2, and β_3 from 6000 iterations after a burn-in phase of 500 are summarized in Figure 1. The posterior means are 4.7, 6.5 and 6.6 for samplers 1,2 and 3, respectively.

The autocorrelation functions for the sampled draws show a relatively fast decline, an indication of the good mixing property of the algorithm. After fifteen lags, the autocorrelations are essentially zero. A similarly fast decline in autocorrelation is also obtained for the simulated draws of the covariance matrix D (Figure 2).

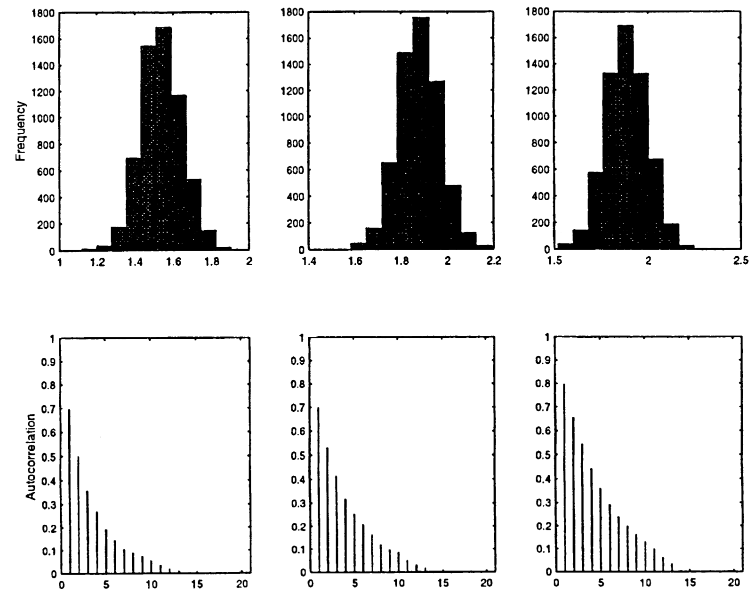

Figure 1: *Posterior distribution of β and autocorrelation functions of sampled draws in air samplers data example.*

The posterior means and standard deviations of D are found to be

$$\begin{pmatrix} 0.308 & 0.025 & -0.084 \\ (0.101) & (0.063) & (0.084) \\ & 0.226 & -0.152 \\ & (0.080) & (0.072) \\ & & 0.425 \\ & & (0.130) \end{pmatrix}$$

There is evidence for substantial extra-Poisson variation, as the diagonal elements of D are relatively large. The covariances do not show a systematic pattern, as $\bar{d}_{12}$ is positive, and $\bar{d}_{13}$ and $\bar{d}_{23}$ are negative. However, the standard deviations of the simulated posterior density are large, and the probability of a zero or positive covariance is small only for d_{23} (0.7 percent). This is also apparent from the Figure 2 where we plot the marginal posterior

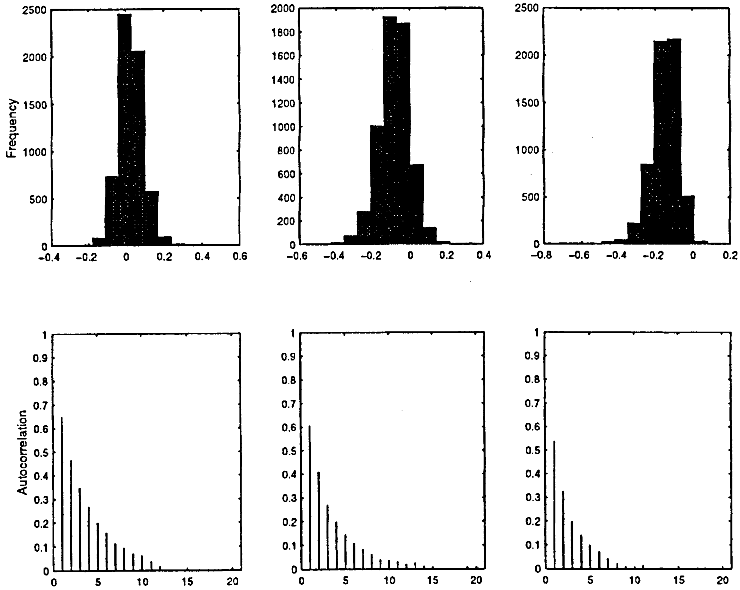

Figure 2: *Posterior distributions and autocorrelation functions from the MCMC output of the covariances* d_{12}, d_{13}, d_{23} *in air samplers data example.*

distributions of the three covariances. One implication of this analysis is that a "one-factor" random effects model would be suspect for these data, as such a model forces the covariances to be positive.

4.2 Seizure counts

The proposed algorithm for computing the posterior density of a multivariate Poisson regression model can also be applied to situations where the dimensionality of J is higher and where covariates are present. Consider the case where j stands for time and the data have a longitudinal structure. For example, Diggle, Liang, and Zeger (1994) provide data on seizure counts (y_{ij}) for each of 59 epileptics over 5 consecutive periods (one observation is eliminated from the dataset because of the "unusually high pre- and post-randomization seizure counts"). Thirty persons are treated with a drug progabide after period one and the following regression model is estimated in order to assess the effectiveness of the treatment:

$$\theta_{ij} = \exp(\beta_1 + \beta_2 x_{ij2} + \beta_3 x_{ij3} + \beta_4 x_{ij4} + e_{ij})$$

where $y_{ij}|\theta_{ij}$ follows a Poisson distribution,

$$x_{ij2} = \begin{cases} 1 & \text{if} \quad \text{visit 1, 2, 3 or 4} \\ 0 & \text{if} \quad \text{baseline} \end{cases}$$

$$x_{ij3} = \begin{cases} 1 & \text{if} \quad \text{treatment group} \\ 0 & \text{if} \quad \text{control} \end{cases}$$

and x_{ij4}, an interaction between x_{ij2} and x_{ij3}. The latent effects are $e_i = (e_{i1}, ..., e_{i5})$ and the prior hyperparameters of β and D (a 5 dimensional matrix) are set to

$$\beta_0 = 0, \; B_0^{-1} = 0.01 I_5; \; \nu_0 = 10, \; R_0 = I_5$$

where $\beta_0 = (\beta_{01}, \beta_{02}, \beta_{03}, \beta_{04})'$.

As the periods are of different length (8 weeks for the base period and 2 weeks each for the post-treatment visits), β_2 accounts for both any genuine difference in the seizure rates before and after treatment for the control group and the effect of the longer base period. For instance, if seizure counts were strictly proportional to the length of the period, one would expect an estimated coefficient of ln(1/4). As the assignment to the drug was randomized, one would expect β_3 to be close to zero. Finally, the coefficient β_4 measures the treatment effect.

With a simulation sample of size 6000 after discarding simulations from a burn-in phase of 500, we obtain marginal posterior distributions of β_1 - β_4 with means and standard deviations of 3.328 (0.126), -1.276 (0.100), -0.011 (0.161), and -0.372 (0.148), respectively.

A box-plot summary of the distributions for the 15 elements of D is presented in Figure 3 along with the autocorrelation functions of the diagonal elements of D. The latter indicate that the sampler is mixing extremely well. The covariances are quite precisely estimated and are all positive, with means between 0.40 and 0.77. The variances tend to be somewhat

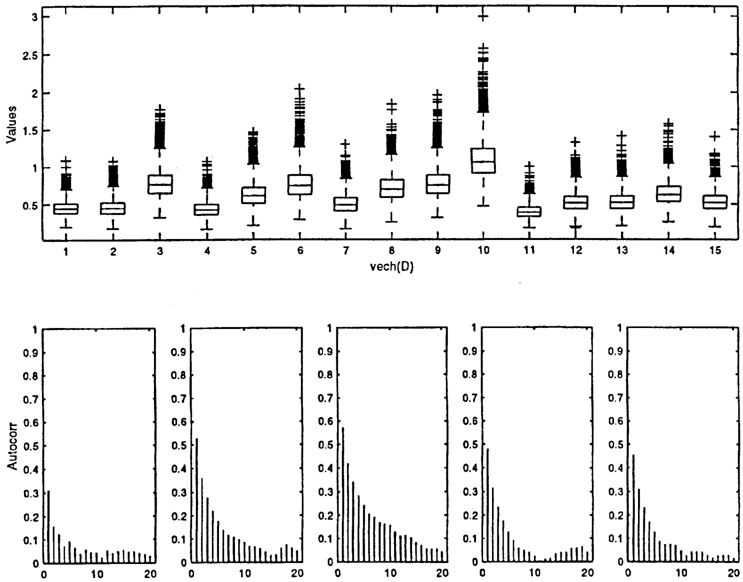

Figure 3: *Posterior box plots of vechD and autocorrelation functions of the diagonal elements of D in epilepsy data example.*

larger than the covariances, in particular d_{44} (column number 10). The evidence appears to suggest only moderate departures from an equi-correlation structure.

We conclude this section by providing a heuristic diagnostic check of the model against the data. Suppose we consider the ability of the model to predict the frequency distribution of the dependent variable. For instance, the outcome "zero" occurs for 23, or 7.9 percent, of all observations. The probability of this outcome from the model depends on β, the latent variable e_{ij} and the covariates x_{it}. One can compute the average predicted probability by integrating $f(0|\beta, e_{ij}, x_{ij})$ over the joint posterior distribution of β and e and over the observed data distribution of x. The average predicted probabilities of other outcomes $y = 1, 2, \ldots$ can be calculated in the same way.

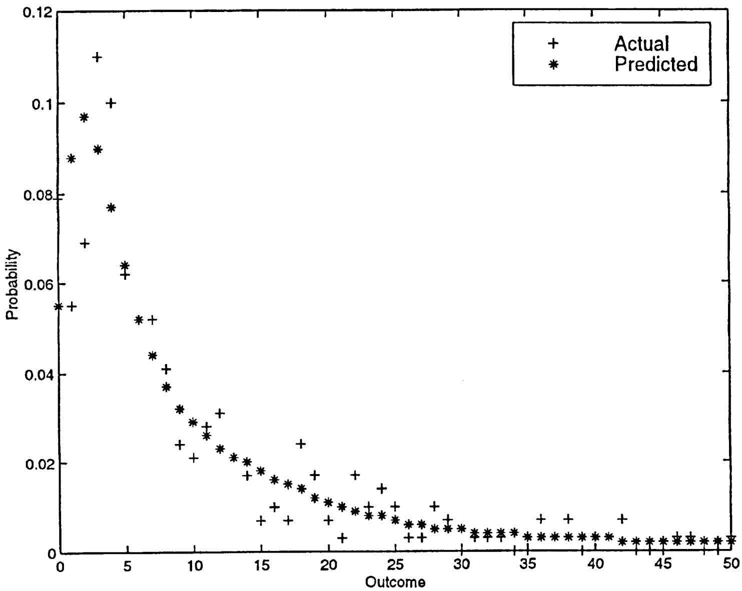

Figure 4: *Relative observed frequencies and average predicted probabilities for epilepsy counts.*

In practice, this approach is very simple to implement as it only requires the output from the MCMC algorithm. A prediction step is included in each iteration using the current values of β and e. The reported average predicted probabilities given in Figure 4 are obtained as grand means over observations and simulations. We find that the predicted distribution traces the data distribution quite closely, although the actual distribution has several non-monotonic parts that the monotonic predicted distribution fails to pick up.

4.3 Number of airline incidents

In this final example we use annual data on the number of airline accidents of sixteen U.S. passenger air carriers between 1957 and 1986, taken from Rose (1990). Carriers with missing observations for some years are excluded from the sample. The accident variable is defined as any operation related occurrence that leads to personal injury or death, or substantial damage to the aircraft. Over the sample, the number of accidents ranges from 0 to 14 with mean 1.7 and variance 4.9.

To model the sixteen counts, let y_{ij} denote the number of accidents in year i for carrier j and let $y_{ij}|\theta_{ij}$ be independently Poisson distributed with conditional mean given by

$$\theta_{ij} = d_{ij} \exp(x'_{ij}\beta + e_{ij}) \ ,$$

where d_{ij} is the total number of departures d_{ij} (in thousands) and the (7×1) vector of covariates x_{ij} includes a constant, the operating margin as a measure of profitability of the airline (OPMARG), the average stage length in thousands of miles (AVSTAGE), the cumulative airline operating experience in billions of aircraft miles (EXPER), the fraction of total departures that are international flights (INTL), an indicator variable for Alaskan carriers, and a linear time trend (see Rose, 1990, for further details).

In this set-up, we allow for contemporaneous correlations between the accident rates of the carriers by assuming that $e_i = (e_{i1}, \ldots, e_{i16})$ are jointly normal distributed with a sixteen dimensional covariance matrix D. We employ the following hyperparameters

$$\beta_0 = 0, \ B_0^{-1} = 0.01 I_7; \ \nu_0 = 32, \ R_0 = I_{16}$$

which implies that the prior mean of the diagonal elements of D is approximately $1/32 = .03$ (indicating small heterogeneity) but with fairly large prior variance (due to the low value of the degree of freedom).

In Table 1 we provide a prior-posterior summary related to β from our MCMC output. The table includes the inefficiency factor (INEFF) (also called the autocorrelation time) in the estimation of the posterior mean of β and defined as $1 + 2\sum_{k=1}^{\infty} \rho(k)$, where $\rho(k)$ is the autocorrelation at lag k for the parameter of interest and the terms in the summation are cut off according to (say) the Parzen window. Each of the inefficiency factors is small indicating that the sampler is mixing well. The quality of the MCMC sampler in the estimation of D is not as easy to summarize given that D contains one hundred and thirty six parameters. To give some idea of the posterior distributions, however, we report in Figure 5 the posterior box plots of the sixteen diagonal elements along with the autocorrelation plots of D_{11}, $D_{9,9}$, $D_{12,12}$ and $D_{16,16}$. We see that the posterior distributions are of the diagonal elements are quite similar with median values ranging from about .07 to .40. Note that the autocorrelations in the sampled output decline quickly indicating once again that the sampler is mixing well.

	Mean	Std dev	Lower	Upper	INEFF
Constant	-4.099	0.105	-4.275	-3.928	4.5011
OPMARG	0.133	0.658	-0.925	1.204	4.4569
AVSTAGE	0.703	0.229	0.328	1.082	5.1250
EXPER	-0.000	0.036	-0.063	0.061	4.1633
INTL	0.459	0.218	0.104	0.824	4.4524
ALASKA	1.005	0.313	0.433	1.481	5.6743
TIME	-0.080	0.006	-0.090	-0.068	4.9788

Table 1: *Posterior summary from the sixteen variate count model fit to airlines data. Results are based on 6000 MCMC draws. "Lower" and "Upper" denote the 5th percentile and the 95th percentile, respectively, and INEFF denotes the inefficiency factor.*

This example provides further evidence of the efficacy of our method in high dimensional count data models that (as far as we are aware) have never before been fit with these many counts or with this level of generality on the correlation structure.

5 Concluding Remarks

The estimation framework developed here can easily be extended to deal with variants of the model discussed above. One example is truncated data. For example, in the analysis of park visitor data one would normally not have information on individuals that have not visited the park and therefore the "zero visits" outcome must be precluded. To deal with this situation one can respecify the conditional Poisson specification for the i, jth count as

$$f(y_{ij}|\beta_j, e_{ij})/(1 - f(0|\beta_j, e_{ij})),\ y_{ij} = 1, 2, 3 \ldots$$

where, as before, f is the Poisson mass function with mean $\theta_{ij} = \exp(x_{ij}'\beta_j + e_{ij})$. Censoring can be also be taken into account in the same way. For instance, data are occasionally top-

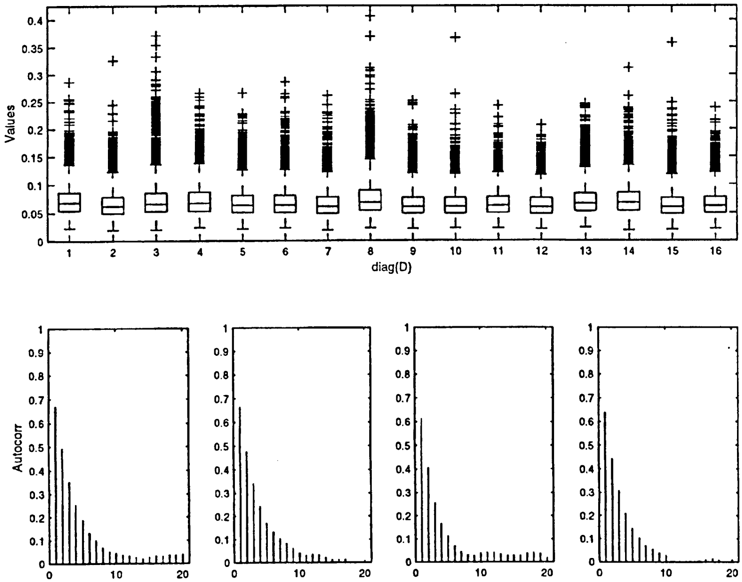

Figure 5: *Posterior box plots of diagD and autocorrelation functions of* D_{11}*,* $D_{9,9}$*,* $D_{12,12}$ *and* $D_{16,16}$ *in airline count data example.*

coded with an open upper category "a or more" counts. The probability mass function is then

$$(1 - F(a|\beta_j, e_{ij}))^{c_{ij}} f(y_{ij}|\beta_j, e_{ij})^{1-c_{ij}},\ y_{ij} = 0, 1, \ldots, a$$

where F is the cumulative distribution function of the Poisson distribution and c_{ij} is an indicator variable that is one if the observation is censored and zero otherwise.

Finally, it is possible to extend the analysis to mixed data consisting of both counts and continuous measurements. The posterior simulation would only require some minor modifications as the data density, conditional on the heterogeneity terms e_{ij}, would now be product of the respective discrete and continuous probability functions.

References

Aitchison, J. and C.H. Ho (1989), "The multivariate Poisson-log normal distribution," *Biometrika*, 76, 643–653.

Cameron, C. and P. Trivedi (1998), *Regression Analysis of Count Data*, Cambridge University Press.

Chib, S. and E. Greenberg (1995), "Understanding the Metropolis-Hastings Algorithm," *American Statistician*, 49, 327–335.

Chib, S., E. Greenberg and R. Winkelmann (1998), "Posterior simulation and bayes factors in panel count data models," *Journal of Econometrics*, 86, 33–54.

Diggle, P.J., K.-Y. Liang and S.L. Zeger (1994), *Analysis of Longitudinal Data*, Oxford University Press.

Gurmu, S. and J. Elder (1998), "Estimation of multivariate count regression models with applications to health care utilization," mimeo, University of Virginia.

Hastings, W.K. (1970), "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika* 57, 97-109.

Hausman, J., B.H. Hall, and Z. Griliches (1984), "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," *Econometrica*, 52, 909–938.

Hinde, L. (1982), "Compound Poisson regression models," in: R. Gilchrist (ed.) *Glim 1982: Proceedings of the international conference on generalized linear models*, Springer: Berlin.

Jung, R.C. and R. Winkelmann (1993), "Two aspects of labor mobility: a bivariate Poisson regression approach," *Empirical Economics* 18, 543–556.

Mayer, W.L. and W.F. Chappell (1992), "Determinants of entry and exit: An application of the compounded bivariate Poisson distribution to U.S. industries, 1972-1977," *Southern Economic Journal*, 58, 770–78.

Rose, N.L. (1990), "Profitability and product quality: Economic determinants of airline safety performance," *Journal of Political Economy*, 98, 944-964.

Winkelmann, R. (1997), *Econometric Analysis of Count Data*, Heidelberg, Springer.

LIST OF DISCUSSION PAPERS

No. 9311 Optimal Long-Run Business Immigration Under Differential Savings Functions, by A. E. Woodfield and K-L. Shea.

No. 9312 The Welfare Cost of Taxation in New Zealand Following Major Tax Reforms, by P. McKeown and A. Woodfield.

No. 9313 The Power of the Goldfeld-Quandt Test when the errors are autocorrelated, by J.P. Small and R.J. Dennis.

No. 9314 The Nucleolus Strikes Back, by M. Carter and P. Walker.

No. 9315 The Output-Inflation Tradeoff in the United States: New evidence on the New Classical vs. New Keynesian Debate, by Alfred V. Guender.

No. 9401 Insurance Market Equilibrium and the Welfare Costs of Gender-Neutral Insurance Pricing under Alternative Regulatory Regimes by Alan E. Woodfield.

No. 9402 Labour Market Signalling and the Welfare Costs of Regulated Insurance Market Equilibria under Gender-neutral Pricing, by Alan E. Woodfield.

No. 9403 The New Classical Vs The New Keynesian debate On The Output – Inflation tradeoff: Evidence From Four industrialized Countries, by Alfred V. Guender.

No. 9404 Yield Spreads & Real Economic Activity: The Case of New Zealand & Australia, by Alfred V. Guender and Mathias Moersch.

No. 9405 Periodic Integration & cointegration with applications to the New Zealand Aggregate Consumption Function, by Robin Harrison and Aaron Smith.

No. 9406 Linear Programming with Mathematica, by Michael Carter.

No. 9407 Are People Really Risk Seeking When Facing Losses? by John Fountain, Michael McCosker & Dean Morris.

No. 9501 Pricing Internet: The New Zealand Experience by Michael Carter and Graeme Guthrie.

No. 9502 Long term and Short Term Aggregate Uncertainty and the Effect on Real Output, by Alfred V. Guender and Robin Young.

No. 9503 Framing and Incentive Effects on Risk Attitudes When Facing Uncertain Losses, by John Fountain, Michael McCosker and Bruce Macfarlane.

No. 9504 An Outline of the History of Game Theory, by Paul Walker.

No. 9505 A Drunk, Her Dog and a Boyfriend: An Illustration of Multiple Cointegration and Error Correction, by Aaron Smith and Robin Harrison.

No. 9506 Optimal Markup Responses to Trade Policy Reform in a Small, Open Economy: Evidence from New Zealand, by Liliana Winkelmann and Rainer Winkelmann.

No. 9507 Unemployment: Where Does It Hurt? by Liliana Winkelmann and Rainer Winkelmann.

No. 9508 Apprenticeship And After: Does It Really Matter? by Rainer Winkelmann.

No. 9601 Another Look at Work Contracts and Absenteeism, by Rainer Winkelmann.

No. 9602 Markov Chain Monte Carlo Analysis of Underreported Count Data With an Application to Worker Absenteeism, by Rainer Winkelmann.

No. 9603 Count Data Models With Selectivity, by Rainer Winkelmann.

No. 9604 Optimal Revenue Smoothing: The Case of New Zealand, by Alfred V. Guender and Kirdan Lees.

No. 9605 Posterior Simulation and Model Choice in Longitudinal Generalized Linear Models, by Siddhartha Chib, Edward Greenberg, Rainer Winkelmann.

No. 9606 Getting It Right: Superannuation and Savings in the U.S.A., by Philip Meguire.

No. 9607 Interconnection in Network Industries, by Michael Carter and Julian Wright.

* Copies of these Discussion Papers may be obtained for $4 (including postage, price changes occasionally) each by writing to the Secretary, Department of Economics, University of Canterbury, Christchurch, New Zealand. A list of the Discussion Papers prior to 1993 is available on request.

This paper is circulated for discussion and comments. It should not be quoted without the prior approval of the author. It reflects the views of the author who is responsible for the facts and accuracy of the data presented. Responsibility for the application of material to specific cases, however, lies with any user of the paper and no responsibility in such cases will be attributed to the author or to the University of Canterbury.