MCMC Variable and Link Determination in Generalised Linear Models

Ioannis Ntzoufras [1], Petros Dellaportas[2] and Jonathan J. Forster[3]

May 13, 1999

SUMMARY

The primary aim of this paper is to discuss and present various MCMC strategies for the simultaneous estimation of posterior variable and link probabilities in generalised linear models. We further discuss the selection of other structral properties such as outlier, transformation or distribution identification. We, finally, illustrate and compare various MCMC algorithms in a binomial variable and link selection example.

Keywords: Gibbs Sampler; link selection; logistic regression; pseudopriors; reversible jump.

1 Introduction

In this paper we concentrate in model selection aspects that may appear in generalised linear models. The usual latent model indicator $m \in \mathcal{M}$ is used to represent the general model formulation including error structure, link function and covariate selection. The set of all possible models $\mathcal{M}$ can be written as a product of two subsets $\mathcal{S} \times \{0,1\}^p$, where $\mathcal{S}$ is the set of all structural properties and $\{0,1\}^p$ the set of all possible combinations of covariates included in the model. Therefore, there is a one-to-one transformation $G : \mathcal{M} \rightarrow \mathcal{S} \times \{0,1\}^p$. The set $\mathcal{S}$ can be further analysed to a product of various other sets such as the set of available links $\mathcal{L}$ or the set of error distributions $\mathcal{D}$. Covariates to be selected are denoted by the usual vector of binary indicator parameters γ ($\gamma_j = 1$ indicates that the j term is included in the model while $\gamma_j = 0$ indicates that the j term is excluded from the model). For example, in a model selection problem where we account uncertainty on the error distribution, link function and covariate selection we use the latent variables $(D, L, \gamma) \in \mathcal{D} \times \mathcal{L} \times \{0,1\}^p$.

In generalised linear models interest usually lies in covariate identification. However, the selection of structural properties is closely related with the validity of the model, its sensitivity to extreme or outlying values, and the structure of randomness in the response variable we study. Therefore here we discuss and present some implementation details when we are interested in other modelling aspects. We focus in link selection but the methodology developed for the link identification can be implemented without any further problems to the determination of other structural properties.

[1]Department of Statistics, Athens University of Economics, Greece, email:jbn@stat-athens.aueb.gr

[2]Department of Statistics, Athens University of Economics, Greece, email:petros@aueb.gr

[3]Department of Mathematics, University of Southampton, email:jjf@maths.soton.ac.uk

The following paper is organized into three sections additional to this introductory one. Section 2 gives variable and link selection details. MCMC algorithms are presented together with a straightforward technique to develop 'equivalent' (in a loose sense) prior distributions among different link functions. In Section 3 we discuss various problems appearing in selection of other structural properties of models including the problem of outlier identification, whether we should try to focus on link or transformation selection and the problem of distribution determination. Finally, in Section 4 we present an illustrative example using simultaneous variable and link selection methods.

2 Covariate and Link Function Identification

Variable and link selection breaks into two steps using the latent variables $\boldsymbol{\gamma}$ for covariate selection and L for link identification. Then we may use any of the available MCMC methods for the covariate selection. We may also incorporate both link and variable selection in one reversible jump Metropolis step though it was proved quite inefficient. The linear predictor will slightly change to

$$g_L(\mu_i) = \eta_i^L, \quad \boldsymbol{\eta}^L = \sum_{j \in \mathcal{V}} \gamma_j \boldsymbol{X}_j \boldsymbol{\beta}_{j,(L)} \tag{1}$$

where $\boldsymbol{\beta}_{j,(L)}$ is the parameter vector that corresponds to j term and L link while the vector of the full model for L link, is noted by $\boldsymbol{\beta}_{(L)}$.

2.1 'Equivalent' Priors for Non-canonical Link Functions

An important issue in link selection is the specification of prior distributions that represent the 'equivalent' beliefs for different link functions. Suppose that we use independent prior distributions for each term conditionally on the link function used. Then the prior of type

$$\boldsymbol{\beta}_{j,(L)} | \gamma_j = 1, L \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_j^L), \text{ for all } L \in \mathcal{L} \tag{2}$$

can be utilized.

A simple approach is based on the notion of semiautomatic selection used in SSVS by George and McCulloch (1993). In the simple case of one dimensional regressors we simply use the equation $[Var(\hat{\beta}_j^L)]^{-1}\Sigma_j^L = [Var(\hat{\beta}_j^C)]^{-1}\Sigma_j^C$ resulting in

$$\Sigma_j^L = \frac{Var(\hat{\beta}_j^L)}{Var(\hat{\beta}_j^C)} \Sigma_j^C$$

where Σ_j^C and Σ_j^L are the prior variances for the canonical link function C and any other link function $L \in \mathcal{L} \setminus \{C\}$, while $Var(\hat{\beta}_j^C)$ and $Var(\hat{\beta}_j^L)$ are the variances of the maximum likelihood estimates $\hat{\beta}_j^C$ and $\hat{\beta}_j^L$ of links C and L, respectively.

In more complicated cases we may use multivariate normal distributions and adjust the variance of each parameter and the prior covariances of parameters β_k and β_l using the equation

$$Cov(\beta_k^L, \beta_l^L) = \frac{Cov(\hat{\beta}_k^L, \hat{\beta}_l^L)}{Cov(\hat{\beta}_k^C, \hat{\beta}_l^C)} Cov(\beta_k^C, \beta_l^C).$$

A more sophisticated adjustment can be based on the first order approximation of Taylor expansion. This is given by

$$\eta^L = g_L(\mu) \approx g_L(\mu_0) + (\mu - \mu_0) g'_L(\mu_0) \tag{3}$$

where $L \in \mathcal{L}$ is a link indicator $g_L(\mu)$ is the link function and η^L the linear predictor corresponding to L indicator. The linear function

$$TE_L(\mu_0) = [g_L(\mu_0) - \mu_0 g'_L(\mu_0)] + \mu g'_L(\mu_0)$$

will be called Taylor expansion linear approximation of link function g_L around μ_0. Then, solving the equation

$$\frac{\eta^{L_1} - g_L(\mu_0)}{g'_{L_1}(\mu_0)} = \frac{\eta^{L_2} - g_L(\mu_0)}{g'_{L_2}(\mu_0)} \tag{4}$$

we have

$$\eta^{L_1} = \delta^*_{L_1L_2}(\mu_0)\eta^{L_2} + \delta^{**}_{L_1L_2}(\mu_0) \tag{5}$$

$$\delta^*_{L_1L_2}(\mu_0) = \frac{g'_{L_1}(\mu_0)}{g'_{L_2}(\mu_0)} \tag{6}$$

$$\delta^{**}_{L_1L_2}(\mu_0) = g_{L_1}(\mu_0) - \delta^*_{L_1L_2}(\mu_0)\eta^{L_2} g_{L_2}(\mu_0). \tag{7}$$

The most frequent application of link selection in generalised linear models is considered in binomial models where the most popular links are the logistic, the probit and the complementary log-log. A common choice for μ_0 is $\mu_0 = p_0 = 0.5$ but this approximation is not effective when data contain many values resulting in binomial probabilities close to one or zero. In such cases, a more effective choice is provided by the mean or median probability of available binomial trials. Details for the binomial case are given in Tables 1 and 2.

L	Link	$g(p_0)$	$g'(p_0)$
L_1	Logit	$log[p_0/(1-p_0)]$	$[p_0(1-p_0)]^{-1}$
L_2	Probit	$\Phi^{-1}(p_0)$	$[\varphi(p_0)]^{-1}$
L_3	clog-log	$log[-log(1-p)]$	$-[(1-p_0)log(1-p_0)]^{-1}$
L_4	log-log	$log[-log(p)]$	$[p_0 log(p_0)]^{-1}$

Table 1: Table of Taylor Expansion for Binomial Example.

We suggest that the linear transformation

$$\beta_{0,(L)} = \delta^*_{LL_1}(\mu_0)\beta_{0,(L_1)} + \delta^{**}_{LL_1}(\mu_0) \tag{8}$$

$$\boldsymbol{\beta}^L_j = \delta^*_{LL_1}(\mu_0)\boldsymbol{\beta}^{L_1}_j \tag{9}$$

is a good approximation between the parameters of the two different links since they satisfy (5).

Alternative prior distributions have been proposed by Raftery (1996). We may also use a prior of type $\boldsymbol{\beta}_{(m)}|m \sim N\left(\boldsymbol{\mu}_{\beta_{(m)}}, c^2\boldsymbol{V}_{(m)}\right)$ with covariance matrix proportional to the inverse of the observed information matrix. Similar approaches for the specification of the prior distributions can be used for transformation identification or distribution selection problems.

L	Link	$g(0.5)$	$g'(0.5)$	TE(0.5)	$\delta^*_{LL_1}$	$\delta^{**}_{LL_1}$
L_1	Logit	0.000	4.000	$4p-2$	1.000	0.000
L_2	Probit	0.000	2.507	$2.507p-1.253$	0.627	0.000
L_3	clog-log	-0.367	2.885	$2.885p-1.809$	0.721	-0.367
L_4	log-log	-0.367	-2.885	$-2.885p+1.076$	-0.721	-0.367

Table 2: Table of Taylor Expansion (TE) for Binomial Example; $p_0 = 0.5$.

2.2 Reversible Jump Link Selection for Given Covariate Structure

The Metropolis step for link selection involves switching link functions while the dimension of the two models remain unchanged and equal to $d(\gamma)$. Two types of reversible jump are appropriate for link selection. The first is to use a suitable transformation without generating any additional terms. The identity transformation may also be used but in some cases (where posterior distribution of parameters in different links are quite distinct) it turned out to be highly inefficient. The second reversible jump type involves a Metropolised Carlin and Chib step where all the parameters are generated from appropriate proposals (see Dellaportas *et al.*, 1998). For more details on the general case of reversible jump see Green (1995)

The reversible jump link step with transformation has acceptance probability

$$\alpha = min\left(1, \frac{f(\boldsymbol{y}|\boldsymbol{\beta}'_{(\gamma,L')},\gamma,L')f(\boldsymbol{\beta}'_{(\gamma,L')}|\gamma,L')f(\gamma,L')j(L',L)}{f(\boldsymbol{y}|\boldsymbol{\beta}_{(\gamma,L)},\gamma,L)f(\boldsymbol{\beta}_{(\gamma,L)}|\gamma,L)f(\gamma,L)j(L,L')}\left|\frac{\partial h_{L,L'}(\boldsymbol{\beta}_{(\gamma,L)})}{\partial(\boldsymbol{\beta}_{(\gamma,L)})}\right|\right) \tag{10}$$

with $h_{L,L'}(\boldsymbol{\beta}_{(\gamma,L)}) = \boldsymbol{\beta}'_{(\gamma,L')}$.

A simple but well performed transformation can be achieved by using the first two terms of the Taylor expansion given by (3). The simple linear transformation given by equations (8) and (9) can be used to assist the chain to jump easily from one link to another. This transformation has the property that the expected values μ_i remain approximately equal among different links since

$$\begin{aligned}
\boldsymbol{\eta}^{L'} &= \beta_{0,(L')} + \sum_{j\in\mathcal{V}\setminus\{0\}} \gamma_j \boldsymbol{X}_j \boldsymbol{\beta}_{j,(L')} \\
&= \delta^*_{L'L}(\mu_0)\beta_{0,(L)} + \delta^{**}_{L'L}(\mu_0) + \delta^*_{L'L}(\mu_0)\sum_{j\in\mathcal{V}\setminus\{0\}} \gamma_j \boldsymbol{X}_j \boldsymbol{\beta}_{j,(L)} \\
&= \delta^*_{L'L}(\mu_0)\boldsymbol{\eta}^L + \delta^{**}_{L'L}(\mu_0),
\end{aligned}$$

which is the result we get when we equate the Taylor approximated expected values for two different links; see (4) and (5).

In general, we may use the value of $p_0 = 0.5$ but in some situations may not be efficient (for example when all the probabilities are small). A better choice for p_0 may be the average probability of the binomial trials. This linear transformation improves the reversible jump step efficiently without having to calculate a complicated Jacobian but the simple $|\delta^{*d(\gamma)}_{LL'}|$.

An alternative approach is to consider a Metropolised Carlin and Chib step and generate all model parameters of the new proposed link from an appropriate proposal distribution. In such case the Metropolis step is given by

$$\alpha = min\left(1, \frac{f(\boldsymbol{y}|\boldsymbol{\beta}'_{(\boldsymbol{\gamma},L')}, \gamma, L')f(\boldsymbol{\beta}'_{(\boldsymbol{\gamma},L')}|\gamma, L')f(\gamma, L')j(L', L)q_L(\boldsymbol{\beta}_{(\boldsymbol{\gamma},L)})}{f(\boldsymbol{y}|\boldsymbol{\beta}_{(\boldsymbol{\gamma},L)}, \gamma, L)f(\boldsymbol{\beta}_{(\boldsymbol{\gamma},L)}|\gamma, L)f(\gamma, L)j(L, L')q_{L'}(\boldsymbol{\beta}'_{(\boldsymbol{\gamma},L')})}\right). \quad (11)$$

Note that the Jacobian here is equal to one since $h_{m,m'}(\boldsymbol{\beta}_{(\boldsymbol{\gamma},L)}, \boldsymbol{\beta}'_{(\boldsymbol{\gamma},L')}) = (\boldsymbol{\beta}'_{(\boldsymbol{\gamma},L')}, \boldsymbol{\beta}_{(\boldsymbol{\gamma},L)})$. The proposal distributions can have the form $q_L(\boldsymbol{\beta}_{(\boldsymbol{\gamma},L)}) = N(\bar{\boldsymbol{\mu}}_{\beta_{(\gamma,L)}}, \boldsymbol{S}_{(\boldsymbol{\gamma},L)})$, where $\bar{\boldsymbol{\mu}}_{\beta_{(\gamma,L)}}$ and $\boldsymbol{S}_{(\boldsymbol{\gamma},L)}$ can be taken by a pilot chain for the model with covariates specified by $\boldsymbol{\gamma}$ and link function specified by L.

Finally, a reversible jump algorithm for variable and link selection using the Dellaportas and Forster (1999) approach in variable selection and one of the above reversible jumps for link selection can be summarized by the following steps:

1. Generate $\boldsymbol{\beta}_{(\boldsymbol{\gamma},L)}$ from the full conditional posterior distribution $f(\boldsymbol{\beta}_{(\boldsymbol{\gamma},L)}|\gamma, L, \boldsymbol{y})$.

2. Select a candidate term j to change with probability $1/p$ and propose to change from γ_j to $1 - \gamma_j$ with probability one.

3a. If $\gamma_j = 0$ then
[i] Generate the additional parameters $\boldsymbol{\beta}'_j$ from the proposal density $q_j(\boldsymbol{\beta}'_j)$.
[ii] Set $\boldsymbol{\beta}'_{(\gamma',L)} = [\boldsymbol{\beta}_{(\gamma,L)}, \boldsymbol{\beta}'_j]$.
[iii] Accept the proposed move with probability

$$\alpha = min\left(1, \frac{f(\boldsymbol{y}|\gamma', L, \boldsymbol{\beta}'_{(\gamma',L)})f(\boldsymbol{\beta}'_{(\gamma',L)}|\gamma', L)f(\gamma', L)}{f(\boldsymbol{y}|\gamma, L, \boldsymbol{\beta}_{(\boldsymbol{\gamma},L)})f(\boldsymbol{\beta}_{(\boldsymbol{\gamma},L)}|\gamma, L)f(\gamma, L)q_j(\boldsymbol{\beta}'_j)}\right). \quad (12)$$

3b. If $\gamma_j = 1$ then
[i] Set $\boldsymbol{\beta}'_{(\gamma',L)}$ equal to the parameters of $\boldsymbol{\beta}_{(\gamma,L)}$ removing the parameters that correspond to j term.
[ii] Accept the proposed move with probability

$$\alpha = min\left(1, \frac{f(\boldsymbol{y}|\gamma', \boldsymbol{\beta}'_{(\gamma',L)})f(\boldsymbol{\beta}'_{(\gamma',L)}|\gamma', L)f(\gamma', L)q_j(\boldsymbol{\beta}_j)}{f(\boldsymbol{y}|\gamma, \boldsymbol{\beta}_{(\boldsymbol{\gamma},L)})f(\boldsymbol{\beta}_{(\boldsymbol{\gamma},L)}|\gamma, L)f(\gamma, L)}\right). \quad (13)$$

4a. If we adopt the approach of transformation then
[i] Propose a new link L' with probability $j(L, L')$.
[ii] Calculate $\boldsymbol{\beta}'_{(\gamma,L')}$ using equations based on Taylor expansion given by (8) and (9).
[iii] Accept the proposed move with probability given by (10).

4b. If we adopt the alternative Metropolised Carlin and Chib step for the link selection then we have
[i] Propose a new link L' with probability $j(L, L')$.
[ii] Propose $\boldsymbol{\beta}'_{(\gamma,L')}$ from a proposal density $q_{L'}(\boldsymbol{\beta}_{(\boldsymbol{\gamma},L')})$.
[iii] Accept the proposed move with probability given by (11).

2.3 Gibbs Variable and Link Selection for Generalised Linear Models

This section demonstrates how we can extend Gibbs variable selection methods to other selection problems such as link selection. Similar methodology may be followed for distribution selection problems.

The first approach in applying Gibbs sampling for variable and link selection problems is to generalise Gibbs variable selection, introduced by Dellaportas *et al.* (1998), or SSVS, introduced by George and McCulloch (1993). We use the same notation as in Gibbs variable selection and therefore $\boldsymbol{\beta}$ denotes the common parameter vector of the full model for any link. Hence the linear predictor (1) is simplified to

$$g_L(\mu_i) = \eta_i^L, \quad \boldsymbol{\eta}^L = \sum_{j \in \mathcal{V}} \gamma_j \boldsymbol{X}_j \boldsymbol{\beta}_j.$$

We consider a partition of $\boldsymbol{\beta}$ into $(\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \boldsymbol{\beta}_{\setminus\boldsymbol{\gamma}})$ corresponding to those components of $\boldsymbol{\beta}$ which are included ($\gamma_j = 1$) or not included ($\gamma_j = 0$) in the model, then the prior $f(\boldsymbol{\beta}|\gamma)$ may be partitioned into model prior $f(\boldsymbol{\beta}_{\boldsymbol{\gamma}}|\gamma)$ and pseudoprior $f(\boldsymbol{\beta}_{\setminus\boldsymbol{\gamma}}|\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \gamma)$.

The resulting posterior conditional distributions for Gibbs variable and link selection are given by

$$\begin{aligned} f(\boldsymbol{\beta}_{(\boldsymbol{\gamma})}|\boldsymbol{\beta}_{(\setminus\boldsymbol{\gamma})}, \boldsymbol{\gamma}, L, \boldsymbol{y}) &\propto f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}, L) f(\boldsymbol{\beta}_{(\boldsymbol{\gamma})}|\boldsymbol{\gamma}, L) f(\boldsymbol{\beta}_{(\setminus\boldsymbol{\gamma})}|\boldsymbol{\beta}_{(\boldsymbol{\gamma})}, \boldsymbol{\gamma}, L) \quad (14) \\ f(\boldsymbol{\beta}_{(\setminus\boldsymbol{\gamma})}|\boldsymbol{\beta}_{(\boldsymbol{\gamma})}, \boldsymbol{\gamma}, L, \boldsymbol{y}) &\propto f(\boldsymbol{\beta}_{(\setminus\boldsymbol{\gamma})}|\boldsymbol{\beta}_{(\boldsymbol{\gamma})}, \boldsymbol{\gamma}, L) \quad (15) \end{aligned}$$

and

$$O_j = \frac{f(\gamma_j = 1|\boldsymbol{\beta}, \boldsymbol{\gamma}_{\setminus j}, L, \boldsymbol{y})}{f(\gamma_j = 0|\boldsymbol{\beta}, \boldsymbol{\gamma}_{\setminus j}, L, \boldsymbol{y})} = \frac{f(\boldsymbol{y}|\boldsymbol{\beta}, \gamma_j = 1, \boldsymbol{\gamma}_{\setminus j}, L)}{f(\boldsymbol{y}|\boldsymbol{\beta}, \gamma_j = 0, \boldsymbol{\gamma}_{\setminus j}, L)} \frac{f(\boldsymbol{\beta}|\gamma_j = 1, \boldsymbol{\gamma}_{\setminus j}, L)}{f(\boldsymbol{\beta}|\gamma_j = 0, \boldsymbol{\gamma}_{\setminus j}, L)} \frac{f(\gamma_j = 1, \boldsymbol{\gamma}_{\setminus j}, L)}{f(\gamma_j = 0, \boldsymbol{\gamma}_{\setminus j}, L)} \quad (16)$$

while the conditional posterior for the link selection step is given by

$$f(L|\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}, L) f(\boldsymbol{\beta}|\boldsymbol{\gamma}, L) f(\boldsymbol{\gamma}, L)}{\sum_{L' \in \mathcal{L}} f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}, L') f(\boldsymbol{\beta}|\boldsymbol{\gamma}, L') f(\boldsymbol{\gamma}, L')} \quad (17)$$

where $f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}, L)$ is the likelihood for the L link.

The main drawback of the above chain is that when the distance of the posterior distributions of the parameters for different links is large then the part of the prior referring to the coefficients outside the current model will slow down the link jumps. A proposed modification is to use of same pseudoprior for all links resulting to their elimination from the link selection step. This will considerably slow down the variable selection step, since proposed values are not optimal, making the Markov chain inefficient. For these reason we shall develop other Gibbs based samplers for simultaneous variable and link selection.

Similarly SSVS can be extended to stochastic search variable and link selection. This extension will be given by posterior densities

$$f(\boldsymbol{\beta}_j|\boldsymbol{\beta}_{\setminus j}, \boldsymbol{\gamma}, L, \boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}, L) f(\boldsymbol{\beta}_j|\gamma_j, L) \quad (18)$$

and

$$\frac{f(\gamma_j=1|\boldsymbol{\beta},\boldsymbol{\gamma}_{\setminus j},L,\boldsymbol{y})}{f(\gamma_j=0|\boldsymbol{\beta},\boldsymbol{\gamma}_{\setminus j},L,\boldsymbol{y})}=\frac{f(\boldsymbol{\beta}|\gamma_j=1,\boldsymbol{\gamma}_{\setminus j},L)}{f(\boldsymbol{\beta}|\gamma_j=0,\boldsymbol{\gamma}_{\setminus j},L)}\frac{f(\gamma_j=1,\boldsymbol{\gamma}_{\setminus j},L)}{f(\gamma_j=0,\boldsymbol{\gamma}_{\setminus j},L)} \tag{19}$$

for the variable selection step while the link step is given by equation (17). Stochastic search variable and link selection demonstrates convergence problems similar to Gibbs variable and link selection sampler especially when the parameters of different links are distinct.

A possible solution to the above problem is given by combining Gibbs variable selection for covariate identification and Carlin and Chib (1995) approach for link selection. Instead of $\boldsymbol{\beta}$ for all link functions we use one vector $\boldsymbol{\beta}_{(L)}$ for each $L \in \mathcal{L}$. Therefore we need to define $f(\boldsymbol{\beta}_{(L')}|\boldsymbol{\gamma},L)$ for all $L, L' \in \mathcal{L}$. When $L'=L$ then the parameter vector $\boldsymbol{\beta}_{(L)}$ is generated from $f(\boldsymbol{\beta}_{(L)}|\boldsymbol{\gamma},L,\boldsymbol{y})$ which is equivalent to Gibbs variable selection steps given by (14) and (15) adding the link indicator L to all model vectors $\boldsymbol{\beta}$, $\boldsymbol{\beta}_{(\boldsymbol{\gamma})}$ and $\boldsymbol{\beta}_{(\setminus\boldsymbol{\gamma})}$. In the case where $L' \neq L$ we generate $\boldsymbol{\beta}_{(L')}$ from the pseudoprior $f(\boldsymbol{\beta}_{(L')}|\boldsymbol{\gamma},L)$. The variable selection step is similar to (16) while the link selection step is given by the conditional posterior

$$f(L|\{\boldsymbol{\beta}_{L'},\ L'\in\mathcal{V}\},\boldsymbol{\gamma},\boldsymbol{y})=\frac{f(\boldsymbol{y}|\boldsymbol{\beta}_{(L)},\boldsymbol{\gamma},L)f(\boldsymbol{\gamma},L)\prod_{L'\in\mathcal{L}}f(\boldsymbol{\beta}^{(L')}|\boldsymbol{\gamma},L)}{\sum_{\ell\in\mathcal{L}}f(\boldsymbol{y}|\boldsymbol{\beta}_{(\ell)},\boldsymbol{\gamma},\ell)f(\boldsymbol{\gamma},\ell)\prod_{L'\in\mathcal{L}}f(\boldsymbol{\beta}^{(L')}|\boldsymbol{\gamma},\ell)}. \tag{20}$$

The above Gibbs sampler solves the problem appeared in Gibbs variable selection or SSVS extension to link identification problems. The major drawback of this Gibbs sampler is the prerequisite of proposing many parameters from the pseudopriors. This can be solved if we metropolise this step resulting in Metropolised Carlin and Chib link selection step as described in Dellaportas *et al.* (1998). Similar procedures can be used to construct flexible samplers that give accurate results in reasonable time. We can also construct a Gibbs sampler that applies SSVS methodology in variable selection and Carlin and Chib (1995) methodology for link selection as described above.

2.4 Metropolised Gibbs Sampler for Link Selection

Liu (1996) described Metropolised Gibbs sampler, an improoved MCMC algorithm, which is optimal for sampling from discrete distributions. This optimal MCMC algorithm can also be implemented for the link selection samplers. We develop this variant in which we propose a new link $L' \neq L$ from the distribution

$$j(L,L')=\frac{f(\boldsymbol{y}|\boldsymbol{\beta}_{(L')},\boldsymbol{\gamma},L')f(\boldsymbol{\gamma},L')\prod_{L''\in\mathcal{L}}f(\boldsymbol{\beta}_{(L'')}|\boldsymbol{\gamma},L')}{\sum_{\ell\in\mathcal{L}\setminus\{L\}}f(\boldsymbol{y}|\boldsymbol{\beta}_{(\ell)},\boldsymbol{\gamma},\ell)f(\boldsymbol{\gamma},\ell)\prod_{L''\in\mathcal{L}}f(\boldsymbol{\beta}_{(L'')}|\boldsymbol{\gamma},\ell)}$$

and accept the proposed term with acceptance probability

$$\alpha=min\left(1,\frac{\sum_{\ell\in\mathcal{L}\setminus\{L\}}f(\boldsymbol{y}|\boldsymbol{\beta}_{(\ell)},\boldsymbol{\gamma},\ell)f(\boldsymbol{\gamma},\ell)\prod_{L''\in\mathcal{L}}f(\boldsymbol{\beta}_{(L'')}|\boldsymbol{\gamma},\ell)}{\sum_{\ell\in\mathcal{L}\setminus\{L'\}}f(\boldsymbol{y}|\boldsymbol{\beta}_{(\ell)},\boldsymbol{\gamma},\ell)f(\boldsymbol{\gamma},\ell)\prod_{L''\in\mathcal{L}}f(\boldsymbol{\beta}_{(L'')}|\boldsymbol{\gamma},\ell)}\right).$$

Although the above sampler seems complicated, it does not require more computations than the above Gibbs variable and Carlin and Chib link selection MCMC algorithm. On the other hand, it is more complicated than both versions of reversible jump for link selection proposed in Section 2.2. The above sampler is given in a simpler form if we set

$$\Xi(\boldsymbol{\beta}^*, \gamma, \ell) = f(\boldsymbol{y}|\boldsymbol{\beta}_{(\ell)}, \gamma, \ell) f(\gamma, \ell) \prod_{L'' \in \mathcal{L}} f(\boldsymbol{\beta}_{(L'')}|\gamma, \ell),$$

$$\Xi(\boldsymbol{\beta}^*, \gamma) = \sum_{\ell \in \mathcal{L}} \Xi(\boldsymbol{\beta}_{(\ell)}, \gamma, \ell),$$

where $\boldsymbol{\beta}^* = \{\boldsymbol{\beta}_{(L'')},\ L'' \in \mathcal{L}\}$. The proposal is now given by

$$j(L, L') = \frac{\Xi(\boldsymbol{\beta}^*, \gamma, L')}{\Xi(\boldsymbol{\beta}^*, \gamma) - \Xi(\boldsymbol{\beta}^*, \gamma, L)}$$

and accept the proposed term with probability

$$\alpha = min\left(1, \frac{\Xi(\boldsymbol{\beta}^*, \gamma) - \Xi(\boldsymbol{\beta}^*, \gamma, L)}{\Xi(\boldsymbol{\beta}^*, \gamma) - \Xi(\boldsymbol{\beta}^*, \gamma, L')}\right).$$

2.5 Other Approaches in Link Selection

The above link selection methods considered a limited number of distinct link functions. More general link functions can be adopted by considering a family of link function depending on one or more (unknown) parameters. The set of all possible links will depend on the possible values that these parameters can take. For example Mallick and Gelfand (1994) consider mixtures of beta distributions while Basu and Mukhopadhyay (1994) normal scale mixtures. Lang (1997) considers the link function

$$g_\varrho(\mu_i) = m_1(\varrho) F_{-\infty}(\mu_i) + m_2(\varrho) F(\mu_i) + m_3(\varrho) F_\infty(\mu_i)$$

where ϱ is a mixing parameter to be estimated and $m_i(\varrho)$ for $i = 1, 2, 3$ are mixing functions. As mixing functions, he proposed $m_1(\varrho) = exp(-e^{3.5\varrho+2})$ $m_3(\varrho) = exp(-e^{-3.5\varrho+2})$ and $m_2(\varrho) = 1 - m_1(\varrho) - m_3(\varrho)$ while $F_{-\infty}(\mu) = 1 - exp(-e^{\mu})$ (extreme minimum value function), $F_\infty(\mu) = exp(-e^{\mu})$ (extreme maximum value function) and $F(\mu) = e^{\mu}/(1+e^{\mu})$ (logistic function). For the mixing parameter ϱ, a normal prior distribution was suggested. A straightforward modification can be done by

$$g_\varrho(\mu) = m_1 F_{-\infty}(\mu) + m_2 F(\mu) + (1 - m_1 - m_2) F_\infty(\mu)$$

where m_i are now mixing proportions to be estimated. A Dirichlet prior on $[m_1, m_2]$ can be used.

Finally, Albert and Chib (1997) propose to use the family of symmetric links given by

$$\mu_i = g_\varrho^{-1}(\eta_i) = \frac{2}{\varrho} \times \frac{\eta_i^\varrho - (1-\eta_i)^\varrho}{\eta_i^\varrho + (1-\eta_i)^\varrho}$$

where $\varrho = 0.0, 0.4, 1.0$ correspond to the logit, (approximately) probit and linear link respectively. Alternatively for binomial models we may use

$$g_\varrho(p_i) = a_\varrho \left(\frac{p_i}{1-p_i}\right)^\varrho + b_\varrho.$$

The choice of $\varrho = 0$, $a_\varrho = 1$ and $b_\varrho = 0$ corresponds to the logit link. We decided not to proceed in comparative studies since these approaches are of different notion.

3 Determination of Other Structural Properties

3.1 Alternative Procedures for Outlier Identification

The most common method for outlier identification is called variance inflation method and was used by Hoeting *et al.* (1996) for linear models and Albert and Chib (1997) for generalised linear models. An alternative approach is to exclude outliers from the main body of the model formulation and estimate its parameters seperately (see Ntzoufras, 1999).

Both of the above proposed method can be adopted in Poisson or binomial model or other non-normal models. In such cases advanced MCMC algorithms, such as reversible jump or a Gibbs variable setup, should be adopted in order to identify outliers. For example, in order to use the second approach in Poisson models we may adopt the linear predictor

$$\eta_i = log(\lambda_i) = v_i \eta_i + (1 - v_i) log(o_i)$$

or in binomial models we may generally consider

$$p_i = \left\{ g^{-1}(\eta_i) \right\}^{v_i} \times o_i^{1-v_i}$$

where η_i is the linear predictor for i subject, $g^{-1}(x)$ is the inverse of the link function, o_i is an outlier specific parameter with $o_i|\gamma_i = 0 \sim Beta(a_o, b_o)$. We additionally need to define proposal distributions for these parameters when they are not included in the likelihood. Straightforward proposals may be given by $o_i|\gamma_i = 1 \sim G(y_i, 1)$ in the Poisson case and $o_i|\gamma_i = 1 \sim Beta(y_i, N_i - y_i)$ in binomial.

In the case where we simply ignore outliers from the likelihood, the procedure for sampling model parameters and covariate indicators is the same as the simple covariate selection methods but the summations (or products) involved in the likelihood are limited to non-outlying observations. The additional outlier step is given by either Gibbs step with probability $O_i^*/(1 + O_i^*)$ or a Metropolis step with acceptance probability $\alpha = min(1, O_i^{*-\nu_i})$; where

$$O_i^* = \frac{f(v_i = 1 | \boldsymbol{\beta}_{(\boldsymbol{\gamma}, v_i=1, \boldsymbol{v}_{\setminus i})}, \boldsymbol{\gamma}, \boldsymbol{v}_{\setminus i}, \boldsymbol{y})}{f(v_i = 0 | \boldsymbol{\beta}_{(\boldsymbol{\gamma}, v_i=0, \boldsymbol{v}_{\setminus i})}, \boldsymbol{\gamma}, \boldsymbol{v}_{\setminus i}, \boldsymbol{y})}$$

where $\boldsymbol{\beta}_{(\boldsymbol{\gamma}, \boldsymbol{v})}$ is model parameter vector of a model with covariates specified by $\boldsymbol{\gamma}$ and outlier specified by $\boldsymbol{v}$. If a Gibbs variable selection type algorithm is adopted it is natural to assume that the prior distribution of model parameters and the covariate prior probability are independent of the outlier structure resulting in

$$O_i^* = f(y_i | \boldsymbol{\beta}_{(\boldsymbol{\gamma}, v_i=1, \boldsymbol{v}_{\setminus i})}, \boldsymbol{\gamma}, v_i = 1, \boldsymbol{v}_{\setminus i}) \frac{f(v_i = 1, \boldsymbol{v}_{\setminus i})}{f(v_i = 0, \boldsymbol{v}_{\setminus i})}.$$

The above equation can be interpreted as a measure of how close is observation i to the current model formulation. If the mean value of the current model is close to the mode of the likelihood for i observation then we include this term in the model with high probability.

3.2 Link or Transformation Selection?

On many cases we will face the problem whether we should concentrate on the link or transformation selection. Using transformations of the responses in linear models changes the distribution of the original response while link selection does not affect the distribution of the original response but changes the connection between the mean and the covariates.

For example, if we use the same transforming function then we have $E[g(Y)] = \boldsymbol{X\beta}$ in transformation selection while $g[E(Y)] = \boldsymbol{X\beta}$ in link selection. Using the first two terms of the Taylor expansion we get an approximation for $E[g(Y)]$ in transformation problems resulting to

$$E[g(Y)] \approx g[E(Y)] + g''[E(Y)]Var(Y)/2.$$

$$Var[g(Y)] \approx g'[E(Y)]Var(Y).$$

The above approximation implies that transformation can handle even problems with heteroscedastic errors and adopt more complicated distributional structures while the link selection only considers possible different connections between the mean of the response and the covariates. An initial exploratory analysis may help to decide what is more appropriate. In other cases we may incorporate both transformation and link selection in MCMC and let the data decide which is more appropriate. The latter approach may result to selection of complicated models with no practical interpretation.

	Distribution of Y	Linear Predictor η_i	$Var(Y_i)$
No Link or Transformation	Normal	$E(Y_i)$	σ^2
Link	Normal	$g[E(Y_i)]$	σ^2
Transformation	Unknown	$g[E(Y_i)] + g''(E(Y_i))Var(Y_i)/2$	$\sigma^2/g'[E(Y_i)]$

Table 3: Comparison of Link and Transformation Attributes in Normal Regression.

3.3 Distribution Selection

Distribution selection can easily be handled in a similar way as link selection in the case where we consider a discrete number of distributions. Two usual cases of distribution selection problems are the normal/Student distribution and the Poisson/negative binomial.

For the selection between normal and Student's we may simply handle it as a model with Student distribution with the degrees of freedom taken as an extra parameter and then draw inferences from its posterior distribution. The likelihood of the equivalent linear model with Student distribution is given by

$$f(\boldsymbol{y}|\boldsymbol{\eta}, \sigma^2, df) = \frac{\Gamma([df+n]/2)}{\Gamma(df/2)(2\pi\sigma^2)^{n/2}} \left[1 + \frac{1}{df\sigma^2}(\boldsymbol{y}-\boldsymbol{\eta})^T(\boldsymbol{y}-\boldsymbol{\eta})\right]^{-(df+n)/2}$$

where $\boldsymbol{\eta}$ is the usual vector of linear predictors and df are the degrees of freedom. For large df the Student distribution is well approximated by the normal.

The negative binomial, can be written as a mixture Poisson distribution with

$$y_i \sim Poisson(\epsilon_i^{\delta} e^{\eta_i}), \quad \epsilon_i \sim G(\theta, \theta)$$

where δ is an indicator parameter for switching from Poisson ($\delta = 0$) to negative binomial ($\delta = 1$) setup. The parameter θ controls the over-dispersion since $E(Y_i) = e^{\eta_i}$ and $Var(Y_i) = e^{\eta_i} + e^{2\eta_i}/\theta$. Large values of θ imply low over-dispersion. The full likelihood is given by

$$f(\boldsymbol{y}, \boldsymbol{\epsilon}|\boldsymbol{\eta}, \theta, \boldsymbol{\gamma}, \delta) = f(\boldsymbol{y}|\boldsymbol{\eta}, \boldsymbol{\epsilon}, \boldsymbol{\gamma}) f(\boldsymbol{\epsilon}|\theta, \delta).$$

For the canonical link we have

$$f(\boldsymbol{y}|\boldsymbol{\eta}, \boldsymbol{\epsilon}, \boldsymbol{\gamma}) = exp\left\{ -\sum_{i=1}^{n} \epsilon_i^{\delta} e^{\eta_i} + \delta \sum_{i=1}^{n} y_i log(\epsilon_i) + \sum_{i=1}^{n} y_i \eta_i - \sum_{i=1}^{n} log(y_i!) \right\}$$

and

$$f(\boldsymbol{\epsilon}|\theta, \delta = 1) exp\left\{ -\theta \sum_{i=1}^{n} \epsilon_i + (\theta - 1) \sum_{i=1}^{n} log(\epsilon_i) + n\theta log(\theta) - nlog[\Gamma(\theta)] \right\}.$$

The sampling procedure is similar to simple variable selection including some parts for generating from the latent variables ϵ_i and the dispersion parameter from the corresponding posterior distributions or a suitably selected pseudoprior. The distribution selection step can be constructed using similar procedures as described by Dellaportas *et al.* (1998, 1999).

4 Illustrated Example

We consider a dataset analysed by Healy (1988). The data, presented in Table 4, reflect the relationship between the number of survivals, the patient condition (more or less severe) and the received treatment (antitoxin medication or not).

	Antitoxin	Death	Survivals
More Severe	Yes	15	6
	No	22	4
Less Severe	Yes	5	15
	No	7	5

Table 4: Logistic Regression Example: Healy (1988) Dataset.

We consider four possible links: the logit, the complementary log-log, the probit and the log-log link functions. The log-log link was used in order to illustrate convergence difficulties when one link function produces posterior parameter densities far away from the others.

We wish to select one of the twenty possible nested models (five for each link function) with response variable the number of survivals and explanatory factors the patient condition and the received treatment. The full model is given by

$$Y_{il} \sim Bin(N_{il}, p_{il}), \quad g_L(p_{il}) = \mu + a_i + b_l + (ab)_{il}, \quad i, l = 1, 2,$$

where $L \in \{1, 2, 3, 4\}$ is link indicator for the logit, the complementary log-log, the probit and the log-log link functions respectively, while Y_{il}, N_{il} and p_{il} are the number of survivals, the total number of patients and the probability of survival under i level of severity and l treatment respectively; μ, a_i, b_l and $(ab)_{il}$ are the model parameters corresponding to the constant term, i level of severity, l level of treatment, and interaction of i severity and l treatment.

For the model coefficients of the canonical (logit) link we used the prior specification suggested by Dellaportas and Forster (1999) when the usual 'sum-to-zero' constrained model parameters are considered. For the binomial logistic regression models of this example we specify $\Sigma_j^1 = 4 \times 2$ in (2) which is equivalent to the prior used above for log-linear model selection. The prior distributions of the other link functions were defined using the Taylor adjustment around 0.4 which is calculated by

$$\bar{p} = \frac{1}{4} \sum_{i=1}^{2} \sum_{l=1}^{2} \left(\frac{Y_{il}}{N_{il}} \right)$$

where Y_{il} and N_{il} are the number of survivals and the total number of patients under i severity and l treatment respectively.

Details of the methods used are provided in Table 5. For all methods a total of 110,000 iteration were considered discarding the first 10,000 as burn in. Reversible jump and Carlin and Chib variable link modification performed very well and reached convergence very fast (mean batch standard deviation lower than 0.005). The simple extensions of Gibbs variable selection, SSVS and Kuo and Mallick (1998) methods did not perform well and did not reach convergence (mean batch standard deviation of 0.027, 0.053, 0.028 respectively). This is confirmed also by the barplots of Figure 1 where the two reversible jumps and Carlin and Chib variable and link selection (CCVLS) have very smooth distributions while the other samplers demonstrate rough changes. Note that Gibbs variable and link selection (GVLS) using pilot run proposal seems to have reached convergence but this is not the case. The log-log link was not visited at all in the 110,000 iterations examined due to the difficulties discussed in Section 2.3. On the other hand, using proposal centered at zero with pseudoprior variance 100 times smaller than the prior variance worked better but convergence was very slow compared to reversible jump or Carlin and Chib (1995) based schemes.

In the three converged methods eight out of the twenty possible models had posterior probability higher than 5%. All these models include the factor severity (A) while four of them additionally include the treatment factor (B). The model with logit link and the severity factor has the highest probability (about 14%) while the second model is given by log-log link and the severity factor with probability of of about 13%. Note that all these eight models are quite close in terms of posterior probabilities and a selection of a single model is not suggested. In such cases Bayesian model averaging is proposed instead of selecting a single model.

Figures 2 and 3 give comparisons of the the ergodic posterior probabilities of the best eight models for different algorithms. In Figure 2 we clearly see that the reversible jumps and Carlin and Chib variable and link selection are quite close. In Figure 3 we compare the rest of the methods used with the the first reversible jump scheme. Only GVLS with $k = 10$ seems to give close results although demonstrates greater variability. Finally, Figure 4 compares the ergodic probabilities when the adjustment of non-canonical link is made by Taylor expansion around $\bar{p} = 0.4$, Taylor expansion around 0.5 and maximum likelihood standard deviations. Taylor approximation around 0.4 and maximum likelihood estimates adjustment gave similar results in clog-log and log-log link functions while two the Taylor based approximations gave similar results for in the probit link. This may indicates that for the probit link the Taylor expansion round 0.5 is sufficient (probably due to the symmetry of the link) while for the other two (asymmetric) links more complicated adjustments are demanded (based on maximum likelihood estimates or Taylor expansion round $\bar{p}$).

		Algorithm for Selection of			
	Abbreviation	Variable		Link	
1	RJ_1	Simple RJ	(eq. 12-13)	RJ with Transformation	(eq. 10)
		[Transformation: eq. 8 & 9 with $p = \bar{p}$]			
2	RJ_2	Simple RJ	(eq. 12-13)	RJ with Pilot Run Proposals	(eq. 11)
3	CCVLS	GVS	(eq. 16)	Carlin and Chib Type	(eq. 20)
4	GVLS ($k = 10$)	GVS	(eq. 16)	Simple Gibbs Step	(eq. 17)
		[Pseudoprior: $N(0, \Sigma_j/k^2)$ for Each Link]			
5	SSVLS	SSVS	(eq. 19)	Simple Gibbs Step	(eq. 17)
6	KMVLS	KM	(eq. 16) with $f(\boldsymbol{\beta}\|\boldsymbol{\gamma}, L) = f(\boldsymbol{\beta}\|L)$	Simple Gibbs Step	(eq. 17)
7	GVLS	GVS	(eq. 16)	Simple Gibbs Step	(eq. 17)
		[Pseudoprior: Full Model Pilot Run for Each Link]			

Table 5: Variable and Link Algorithms Used in Healy Data.

Model		Posterior Probability			
Factors	Link	RJ_1	RJ_2	CCVLS	GVLS ($k = 10$)
A	logit	0.139	0.138	0.137	0.140
A	log-log	0.130	0.131	0.130	0.134
A	probit	0.127	0.127	0.126	0.126
A+B	logit	0.125	0.125	0.130	0.123
A	cloglog	0.121	0.121	0.122	0.121
A+B	cloglog	0.121	0.121	0.122	0.121
A+B	probit	0.100	0.101	0.103	0.096
A+B	log-log	0.073	0.072	0.077	0.078

Table 6: Link and Variable Selection for Healy Data: Posterior Model Probabilities Higher than 0.05; See Table 5 for Acronyms.

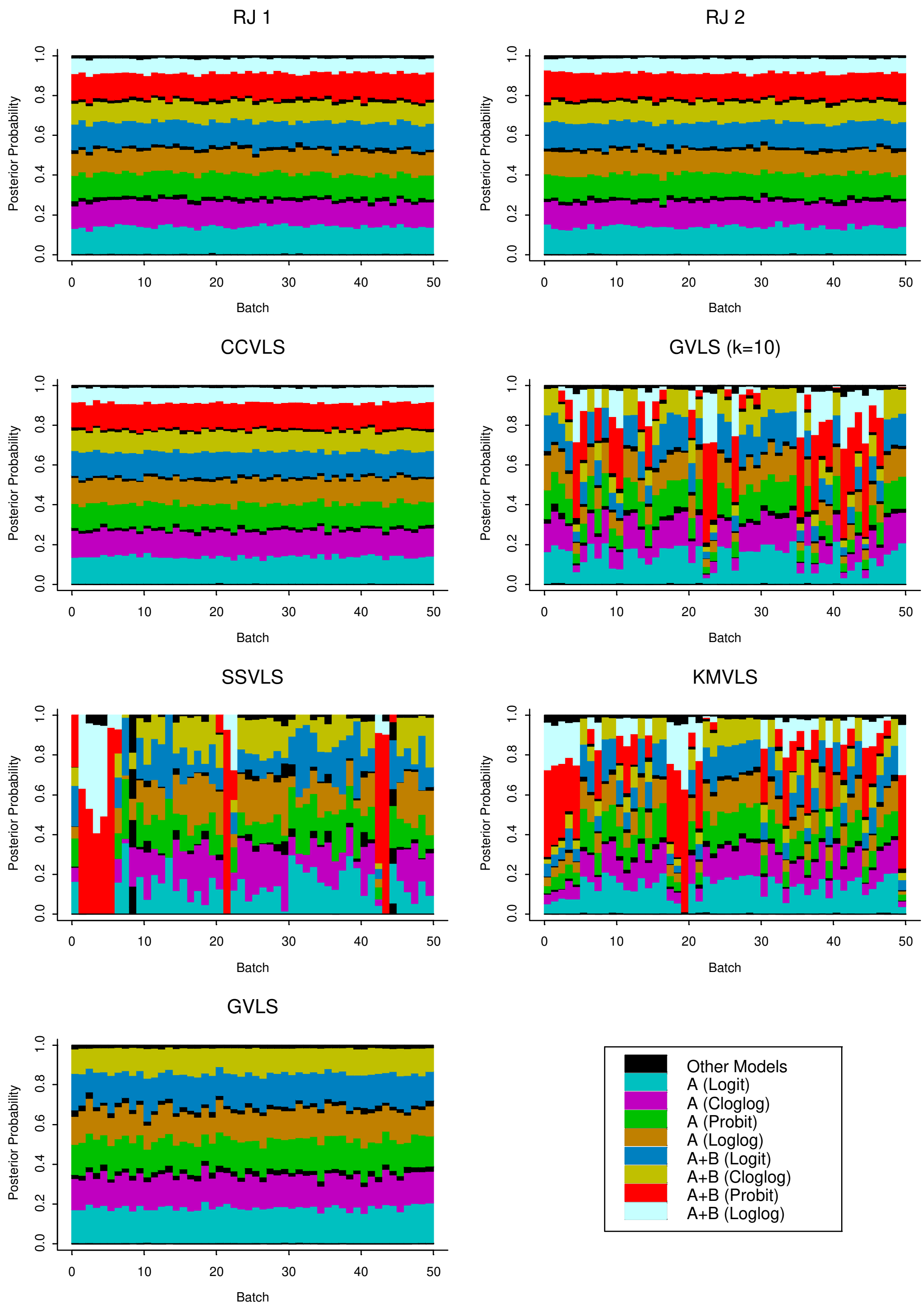

Figure 1: Link and Variable Selection in Healy Data: Batch Posterior Model Probabilities of Different Variable and Link Selection Methods; See Table 5 for Acronyms.

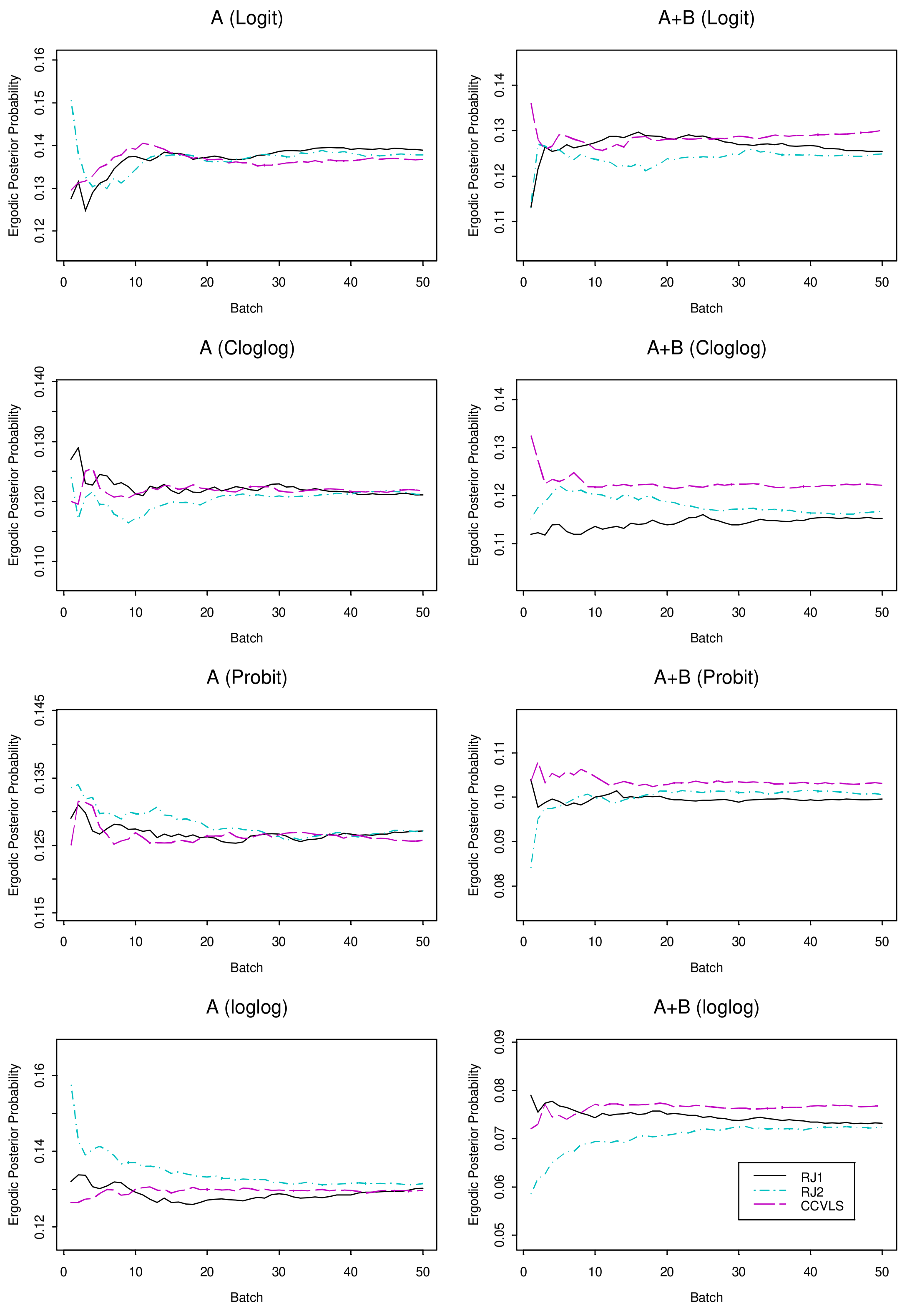

Figure 2: Link and Variable Selection in Healy Data: Ergodic Posterior Model Probabilities of RJ_1, RJ_2 and CCVLS; See Table 5 for Acronyms.

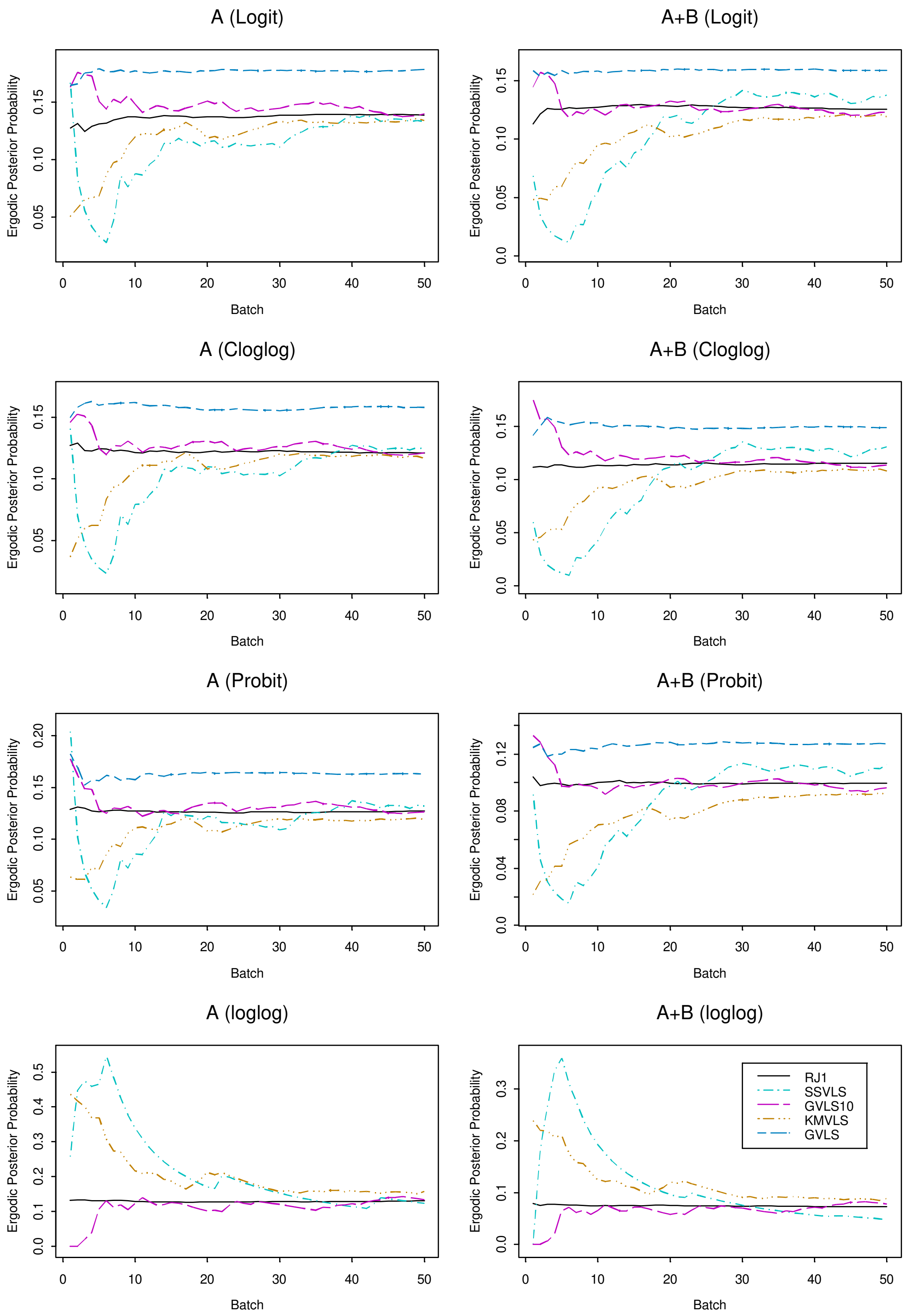

Figure 3: Link and Variable Selection in Healy Data: Ergodic Posterior Model Probabilities Comparison of RJ_1, SSVLS, GVLS (k=10), KMVLs and GVLS; See Table 5 for Acronyms.

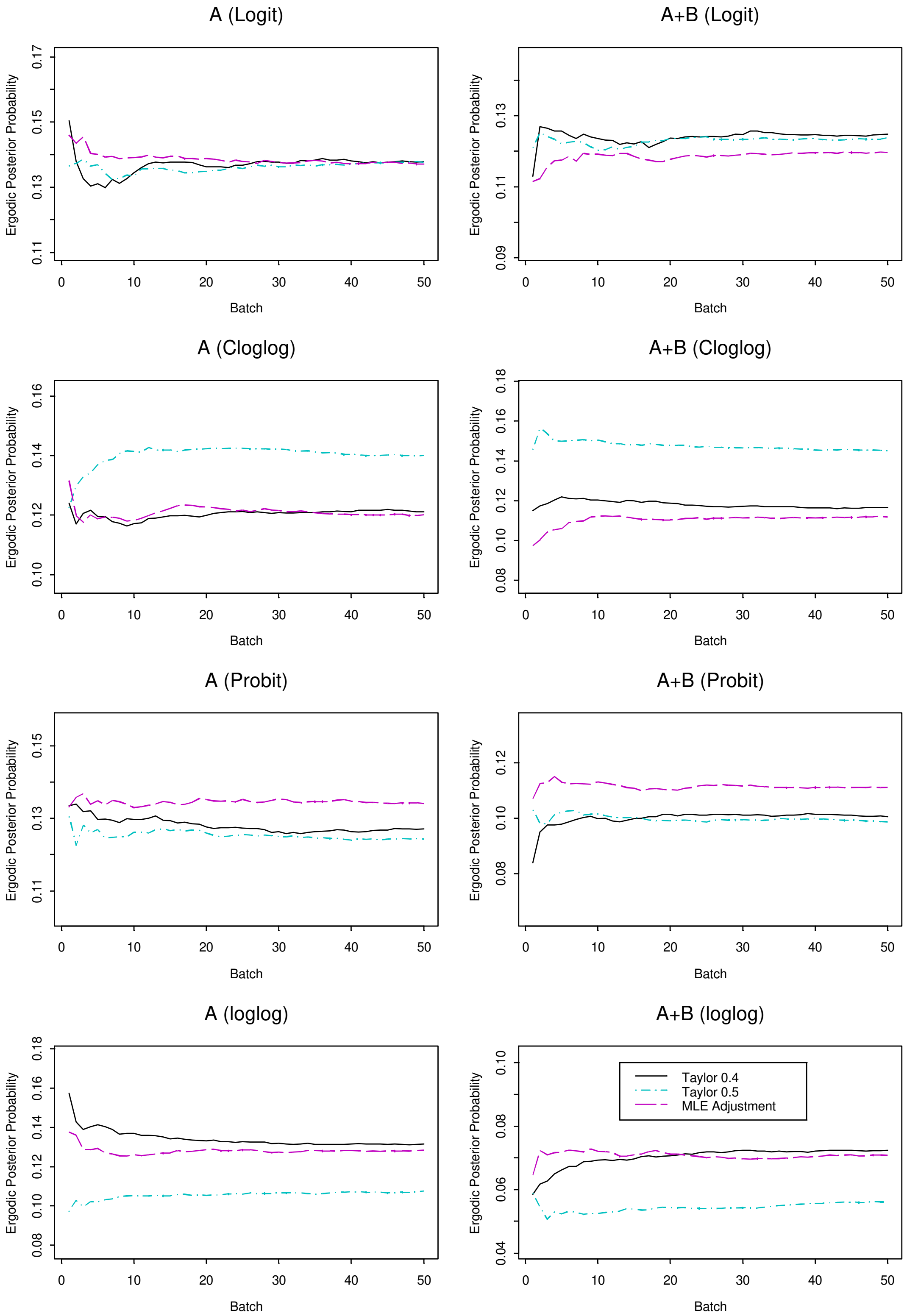

Figure 4: Link and Variable Selection in Healy Data: Comparison of Ergodic Posterior Model Probabilities for Different Prior Distribution Link Adjustment.

	Posterior Probability			
Link	RJ_1	RJ_2	CCVLS	GVLS ($k = 10$)
Logit	0.284	0.283	0.282	0.282
Cloglog	0.253	0.254	0.257	0.251
Probit	0.243	0.244	0.242	0.238
Loglog	0.220	0.220	0.219	0.230

Table 7: Link and Variable Selection in Healy Data: Marginal Posterior Distribution of Link Functions; See Table 5 for Acronyms.

References

[1] Albert, J.H. and Chib, S. (1993). Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association*, **88**, 669–679.

[2] Basu and Mukhopadhyay (1994). Bayesian Analysis of a Random Link Function in Binary Response Regression. *Technical Report*, Department of Mathematical Sciences, University of Arkansas, USA.

[3] Carlin, B.P. and Chib, S. (1995). Bayesian Model Choice via Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society B*, **157**, 473–484.

[4] Dellaportas, P. and Forster, J.J. (1999). Markov Chain Monte Carlo Model Determination for Hierarchical and Graphical Log-linear Models. *Biometrika*, (to appear).

[5] Dellaportas, P., Forster, J.J. and Ntzoufras, I. (1998). On Bayesian Model and Variable Selection Using MCMC. *Technical Report* **40**, Department of Statistics, Athens University of Economics and Business, Greece. (submitted, currently under revision).

[6] Dellaportas, P., Forster, J.J. and Ntzoufras, I. (1999). Bayesian Variable Selection Using the Gibbs Sampler. *Generalized Linear Models: A Bayesian Perspective* (D. K. Dey, S. Ghosh, and B. Mallick, eds.). New York: Marcel Dekker, (to appear).

[7] George, E.I. and McCulloch, R.E. (1993). Variable Selection via Gibbs Sampling. *Journal of the American Statistical Association*, **88**, 881–889.

[8] Green, P. (1995). Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika*, **82**, 711–732.

[9] Healy, M.J.R. (1988). *Glim: An Introduction.* Oxford: University Press.

[10] Hoeting, J.A., Madigan, D. and Raftery, A.E. (1996). A Method for Simultaneous Variable Selection and Outlier Identification in Linear Regression. *Journal of Computational Statistics and Data Analysis*, **22**, 251-270.

[11] Kuo, L. and Mallick, B. (1998). Variable Selection for Regression Models. *Sankhyā B*, **60**, 65–81.

[12] Lang, J.B. (1997). Bayesian Ordinal and Binary Regression Models with a Parametric Family of Mixture Links. *Technical Report* **267**, Department of Statistics and Actuarial Science, University of Iowa, USA.

[13] Liu, J.S. (1996). Peskun's Theorem and a Modified Discrete-State Gibbs Sampler. *Biometrika*, **83**, 681–682.

[14] Mallick, B.K. and Gelfand, A.E. (1994). Generalized Linear Models with Unknown Number of Components. *Biometrika*, **81**, 237–245.

[15] Ntzoufras, I. (1999). Discussion on 'Bayesian Model Averaging and Model Search Strategies'. *Bayesian Statistics 6* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). Oxford: University Press, (to appear).

[16] Raftery, A.E. (1996). Approximate Bayes Factors and Accounting for Model Uncertainty in Generalized Linear Models. *Biometrika*, **83**, 251–266.