Fitting and Comparison of Models for Multivariate Ordinal Outcomes

Ivan Jeliazkov* Jennifer Graves Mark Kutzbach

August 2008

Abstract

In this paper we consider the analysis of models for univariate and multivariate ordinal outcomes in the context of the latent variable inferential framework of Albert and Chib (1993). We review several alternative modeling and identification schemes and evaluate how each aids or hampers estimation by Markov chain Monte Carlo simulation methods. For each identification scheme we also discuss the question of model comparison by marginal likelihoods and Bayes factors. In addition, we develop a simulation-based framework for analyzing covariate effects that can provide interpretability of the results despite the non-linearities in the model and the different identification restrictions that can be implemented. The methods are employed to analyze problems in labor economics (educational attainment), political economy (voter opinions), and health economics (consumers' reliance on alternative sources of medical information).

Keywords: Accept-reject Metropolis-Hastings sampling; data augmentation; discrete data; educational attainment; Gibbs sampling; health information; latent data; limited dependent variable models; marginal likelihood; Markov chain Monte Carlo; survey data; voter opinions.

1 Introduction

This article considers three main inferential problems, namely those of identification, estimation, and model comparison, in the context of models for ordinal outcomes. We exploit the inferential framework of Albert and Chib (1993), which capitalizes on the latent variable representation of binary and categorical response models to simplify the analysis of such problems. In our setting this framework lends itself to efficient fitting by Markov chain Monte Carlo (MCMC) methods—in some instances it allows for direct sampling from known full-conditional distributions, and in others it facilitates the application of versatile simulation techniques such as the Metropolis-Hastings (MH)

*Department of Economics, University of California, Irvine, 3151 Social Science Plaza, Irvine CA 92697-5100. E-mail addresses: ivan@uci.edu, jgraves@uci.edu, and kutzbach@uci.edu. We are much indebted to Siddhartha Chib, William Griffiths, and an anonymous referee for their help with earlier drafts of this paper.

and accept-reject Metropolis-Hastings (ARMH) algorithms (Gelfand and Smith 1990; Metropolis et al. 1953; Hastings 1970; Tierney 1994; Chib and Greenberg 1995). In the ordinal context, we review alternative sets of identification constraints and evaluate how each aids or hampers estimation by MCMC methods. We then consider the issue of model comparison by showing how marginal likelihoods and Bayes factors can be computed in the ordinal data setting using the method presented in Chib (1995) and its extensions developed in Chib and Jeliazkov (2001, 2005). This, for instance, allows for the formal comparison of models with different correlation structures, covariates, or link functions. In addition, we describe a simulation-based approach for calculating the effect of covariates on the outcome, which provides interpretability of the estimates despite the non-linearity of the model and the different identification schemes that can be used to identify the parameters. We apply our methods to three problems in economics involving educational attainment, reliance on health care information sources, and exit poll data on voter opinions about the economy, the war in Iraq, and President George W. Bush's performance in office.

To illustrate the setting, consider a simple univariate case where y_i is a scalar response variable that takes one of the J ordered values, $j = 1, ..., J$, and the index i $(i = 1, ..., n)$ refers to units in the sample (e.g. individuals, families, firms, etc.). The defining feature of ordinal data is that the outcomes are arranged and measured on a monotone scale – e.g. in quantifying survey responses, 1 could be assigned to "very unhappy", 2 to "not too happy", 3 to "happy", and 4 to "very happy"; however, the scale is not assumed to be cardinal, so that differences between categories are not directly comparable. In other words, while the scale tells us that 4 implies more happiness than 2, this does not mean that 4 implies twice as much happiness as 2, or that the difference in happiness between 1 and 3 is the same as that between 2 and 4. Models for ordinal data address these features of the data by postulating a data generating process in which the outcomes can be thought of as arising from an underlying latent variable threshold-crossing framework. In particular, the problem can be motivated by assuming that a continuous latent random variable z_i depends on a k-vector

of covariates $\boldsymbol{x}_i$ through the model

$$z_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \quad i = 1, ..., n, \tag{1}$$

and that the outcome y_i arises according to

$$y_i = j \quad \text{if} \quad \gamma_{j-1} < z_i \leq \gamma_j, \tag{2}$$

where $E(\varepsilon_i|\boldsymbol{x}_i) = 0$ and $-\infty = \gamma_0 < \gamma_1 < ... < \gamma_{J-1} < \gamma_J = \infty$ are cutpoint parameters that determine the discretization of the data into the J ordered categories. Given this representation and a cumulative distribution function (cdf) for ε_i, $F(\varepsilon_i)$, the probability of observing $y_i = j$, conditional on $\boldsymbol{\beta}$ and $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_{J-1})'$, is given by $\Pr(y_i = j|\boldsymbol{\beta}, \boldsymbol{\gamma}) = \Pr(\{\gamma_{j-1} < z_i\} \bigcap \{z_i \leq \gamma_j\}) = \Pr(\{\gamma_{j-1} < \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i\} \bigcap \{\boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i \leq \gamma_j\})$. Letting $A = \{\gamma_{j-1} < \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i\}$ and $B = \{\boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i \leq \gamma_j\}$, from set theory we know that $\Pr(A \bigcap B) = \Pr(A) + \Pr(B) - \Pr(A \bigcup B)$. Therefore, since $\Pr(A) = 1 - F(\gamma_{j-1} - \boldsymbol{x}_i'\boldsymbol{\beta})$, $\Pr(B) = F(\gamma_j - \boldsymbol{x}_i'\boldsymbol{\beta})$, and $\Pr(A \bigcup B) = 1$, we obtain that

$$\Pr(y_i = j|\boldsymbol{\beta}, \boldsymbol{\gamma}) = F\left(\gamma_j - \boldsymbol{x}_i'\boldsymbol{\beta}\right) - F\left(\gamma_{j-1} - \boldsymbol{x}_i'\boldsymbol{\beta}\right). \tag{3}$$

Given $\boldsymbol{y} = (y_1, \ldots, y_n)'$, the likelihood function for the model can be written as

$$f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{i=1}^{n} \prod_{j=1}^{J} \left[F\left(\gamma_j - \boldsymbol{x}_i'\boldsymbol{\beta}\right) - F\left(\gamma_{j-1} - \boldsymbol{x}_i'\boldsymbol{\beta}\right)\right]^{1\{y_i = j\}}, \tag{4}$$

where $1\{y_i = j\}$ is the indicator function of the event $y_i = j$, which takes the value 1 if the event is true and 0 otherwise. Various choices for the cdf $F(\cdot)$ are possible (specific ones will be mentioned below), but practical applications most commonly rely on the Gaussian cdf. For the purposes of illustration, Figure 1 depicts the probabilities of y_i falling in category j as determined by (3) for a four-category setting.

However, both location and scale restrictions are necessary to uniquely identify the parameters of the model. To see this, consider the probabilities in (3) and let $\gamma_j^* = \gamma_j + c$ and $\boldsymbol{x}_i'\boldsymbol{\beta}^* = \boldsymbol{x}_i'\boldsymbol{\beta} + c$ for some constant c (note that the latter is always possible since $\boldsymbol{x}_i$ is assumed to contain a constant term). Then, because $\gamma_j^* - \boldsymbol{x}_i'\boldsymbol{\beta}^* = \gamma_j + c - \boldsymbol{x}_i'\boldsymbol{\beta} - c = \gamma_j - \boldsymbol{x}_i'\boldsymbol{\beta}$, it is straightforward to verify

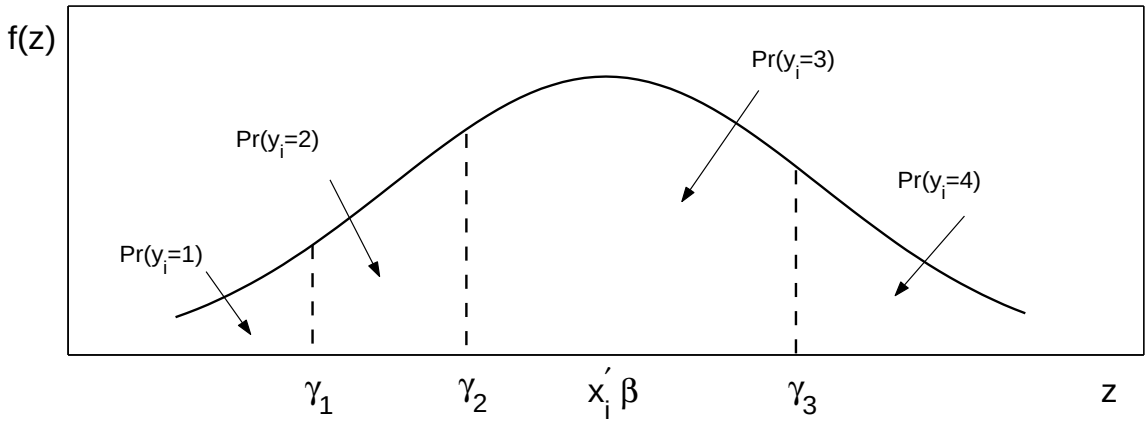

Figure 1: Outcome probabilities for the four-category case given the mean $x_i'\beta$ and the cutpoints γ_1, γ_2, and γ_3.

that $\Pr(y_i = j|\boldsymbol{\beta}, \boldsymbol{\gamma}) = \Pr(y_i = j|\boldsymbol{\beta}^*, \boldsymbol{\gamma}^*)$. This identification problem is usually easily corrected by fixing a cutpoint (in addition to $\gamma_0 = -\infty$ and $\gamma_J = \infty$) — in particular, letting $\gamma_1 = 0$ removes the possibility for shifting the distribution without changing the probability of observing y_i. While in principle it is possible to achieve identification by dropping the intercept instead of fixing a cutpoint, the choice of letting $\gamma_1 = 0$ has the practical benefit of facilitating posterior sampling (since simulating $\boldsymbol{\beta}$ is generally easier than simulating $\boldsymbol{\gamma}$) and also makes the ordinal probit model theoretically consistent with the binary data probit model when there are only two outcome categories (in the binary probit model $y_i = 1$ if $z_i > 0$ and $y_i = 0$ otherwise). For a survey on Bayesian models for ordered categorical data see Liu and Agresti (2005).

The first panel of Figure 2 shows the above considerations regarding location restrictions; the second panel of that figure, however, shows that even if one sets $\gamma_1 = 0$, a second restriction is necessary in order to fix the scale of the latent data that is implied by $F(\cdot)$. Specifically, the second panel of Figure 2 shows that in the absence of additional constraints, one can change the scale of $F(\cdot)$ and simultaneously rescale the mean and the remaining free cutpoints without affecting the probabilities for y_i, implying lack of likelihood identification. This is due to the fact that so far we have only required $F(\cdot)$ to be the cdf of a mean zero distribution, but $F^*(\gamma_j - \boldsymbol{x}_i'\boldsymbol{\beta}) \equiv F\left(\frac{\gamma_j - \boldsymbol{x}_i'\boldsymbol{\beta}}{c}\right)$ is another cdf in the same class that, given mean and cutpoint parameters that are appropriately

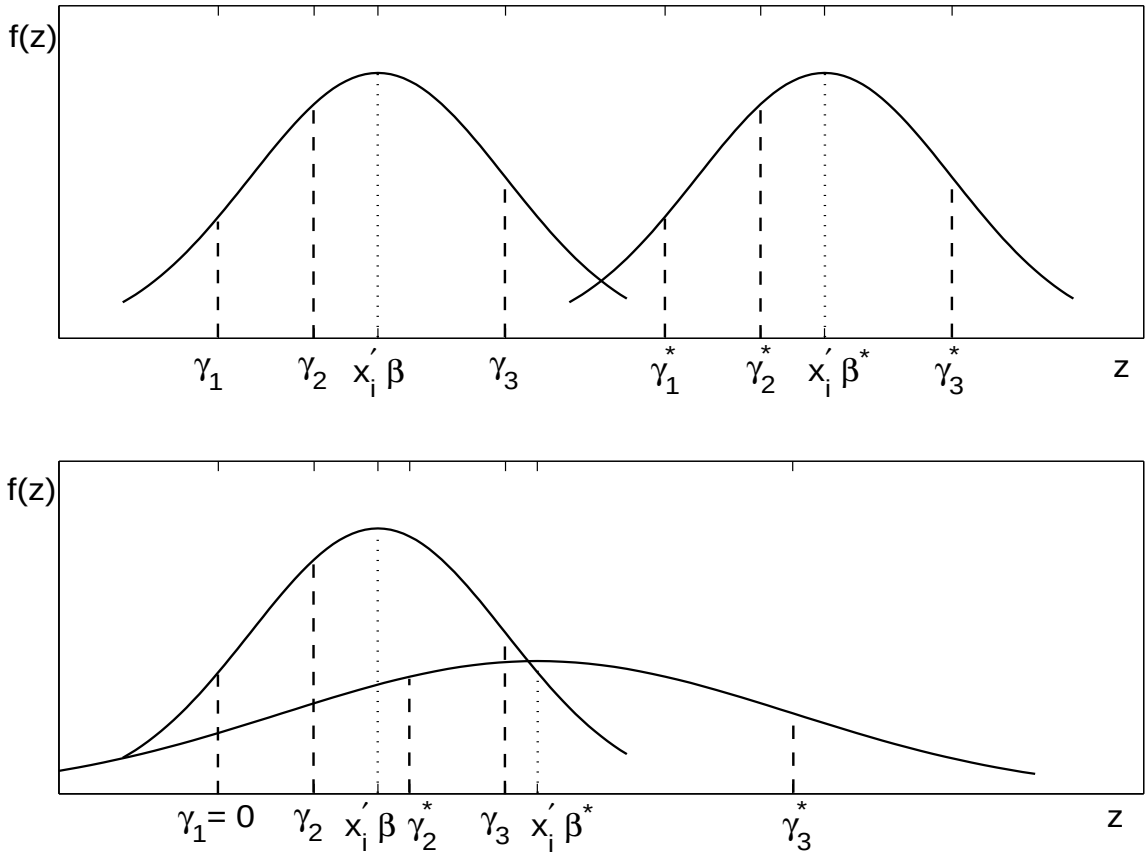

Figure 2: Location and scale constraints are required to identify the outcome probabilities. In the first panel the same probabilities result when the mean and cutpoints are shifted by the same constant. In the second panel, even though $\gamma_1 = 0$, the remaining cutpoints and the mean can be rescaled by a multiplicative constant to produce the same probabilities.

rescaled by a positive constant c, can produce the same probabilities as $F(\cdot)$. The usual approach to achieving identification in this case is to fix the variance of ε. For example, when ε is assumed to be Gaussian, the restriction $var(\varepsilon) = 1$ is usually imposed, leading to an ordered probit model whose link function $F(\cdot)$ is assumed to be the standard normal cdf $\Phi(\cdot)$. Other choices for the link function $F(\cdot)$ include the logistic cdf $(1 + e^{-\varepsilon})^{-1}$ where the variance of ε is given by $\pi^2/3$, the extreme value cdf $e^{-e^{\varepsilon}}$, that implies $var(\varepsilon) = \pi^2/6$, or the t-link model where $F(\cdot)$ is taken to be the standard Student-t cdf with ν degrees of freedom implying $var(\varepsilon) = \nu/(\nu - 2)$. However, in addition to fixing the variance of ε, there are other possible ways to identify the scale of the model. The presence of these alternatives makes it possible to explore various approaches to estimation.

We examine these methods and report on the implementation and performance of the different MCMC estimation algorithms that they determine.

The remainder of this paper is organized as follows. Section 2 discusses a number of identification restrictions together with their corresponding estimation methods in the univariate setting. Section 3 presents the MCMC fitting method for multivariate settings. Section 4 is concerned with model comparison. In Section 5, we show how to compute the effects of covariates on the outcome probability and discuss several extensions of the methods presented in this paper. Section 6 is devoted to the analysis of data on educational attainment, voter opinions, and health information, while brief concluding remarks are presented in Section 7.

2 Ordinal Data Models for Univariate Outcomes

2.1 Estimation Under Traditional Identification Assumptions

Albert and Chib (1993) showed how the latent variable representation of ordinal data models can be exploited for estimating models for ordered data such as the ordinal probit and student-t models. Their idea was to focus not on the posterior distribution of the parameters conditioned on the data but rather on the posterior distribution of the latent data $\boldsymbol{z} = (z_1, ..., z_n)'$ and the parameters conditioned on the data, which is simpler and more tractable to deal with in the context of MCMC methods. In particular, under the probit model assumption that $\varepsilon_i \sim N(0, 1)$, and given the priors $\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_0, \boldsymbol{B}_0)$ and $\pi(\boldsymbol{\gamma}) \propto 1$, the posterior distribution for the latent data and the parameters is given by

$$\begin{aligned} \pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{z}|\boldsymbol{y}) &\propto f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{z})\, \pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{z}) \\ &= f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{z})\, \pi(\boldsymbol{z}|\boldsymbol{\beta})\, \pi(\boldsymbol{\beta})\, \pi(\boldsymbol{\gamma}) \\ &= \left\{ \prod_{i=1}^{n} f(y_i|z_i, \boldsymbol{\gamma}) \right\} \pi(\boldsymbol{z}|\boldsymbol{\beta})\, \pi(\boldsymbol{\beta})\, \pi(\boldsymbol{\gamma}), \end{aligned} \tag{5}$$

where the second line used the decomposition $\pi(\boldsymbol{\beta}, \boldsymbol{z}, \boldsymbol{\gamma}) = \pi(\boldsymbol{z}|\boldsymbol{\beta})\, \pi(\boldsymbol{\beta})\, \pi(\boldsymbol{\gamma})$ that is afforded by prior independence, and the third line used the fact that given the latent z_i and $\boldsymbol{\gamma}$, the observed

y_i is independent of $\boldsymbol{\beta}$ because (2) determines y_i given z_i and $\boldsymbol{\gamma}$ with probability one and that relationship does not involve $\boldsymbol{\beta}$. Specifically, the probability of $y_i = j$ given z_i and $\boldsymbol{\gamma}$ equals 1 when $\gamma_{j-1} < z_i \leq \gamma_j$ and 0 otherwise, so that $f(y_i|z_i, \boldsymbol{\gamma}) = 1\{\gamma_{j-1} < z_i \leq \gamma_j\}$. Also note that $\pi(\boldsymbol{z}|\boldsymbol{\beta})$ can be obtained from (1) and is given by $\pi(\boldsymbol{z}|\boldsymbol{\beta}) = \prod_{i=1}^{n} N(z_i|\boldsymbol{x}_i'\boldsymbol{\beta}, 1)$. With these considerations, the "complete data posterior" in (5) involving the latent data and the parameters is given by

$$\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{z}|\boldsymbol{y}) \propto \left\{ \prod_{i=1}^{n} 1\{\gamma_{j-1} < z_i \leq \gamma_j\} N\left(z_i|\boldsymbol{x}_i'\boldsymbol{\beta}, 1\right) \right\} N(\boldsymbol{\beta}|\boldsymbol{\beta}_0, \boldsymbol{B}_0).$$

Then, upon letting $\boldsymbol{X} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_n)'$, and under the "traditional" identification constraints $\gamma_0 = -\infty$, $\gamma_1 = 0$, $\gamma_J = \infty$, and $var(\varepsilon) = 1$, the posterior distribution of the latent data and the parameters can be sampled by MCMC methods as follows:

Algorithm 1 ***(Albert & Chib 1993) Sampling in the univariate ordered probit model***

1. *For $j = 2, ..., J-1$, sample $\gamma_j|\boldsymbol{z} \sim U(\max\{z_i : y_i = j\}, \min\{z_i : y_i = j+1\})$, i.e., sample γ_j from a uniform distribution bounded between the maximum z_i in category j and the minimum z_i in category $j+1$;*

2. *For $i = 1, ..., n$, sample $z_i|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\gamma} \sim TN_{(\gamma_{j-1}, \gamma_j)}(\boldsymbol{x}_i'\boldsymbol{\beta}, 1)$, where the support for this truncated normal distribution is determined by the cutpoints γ_{j-1} and γ_j associated with $y_i = j$;*

3. *Sample $\boldsymbol{\beta}|\boldsymbol{z} \sim N\left(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{B}}\right)$, where $\hat{\boldsymbol{B}} = (\boldsymbol{B}_0 + \boldsymbol{X}'\boldsymbol{X})^{-1}$ and $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{B}}\left(\boldsymbol{B}_0^{-1}\boldsymbol{\beta}_0 + \boldsymbol{X}'\boldsymbol{z}\right)$, where given $\boldsymbol{z}$, the full-conditional distribution of $\boldsymbol{\beta}$ does not depend on $\boldsymbol{y}$ or $\boldsymbol{\gamma}$.*

This algorithm forms a basis for the fitting of a wide variety of univariate ordinal models because once the probit case can be worked out, a number of other link functions $F(\cdot)$ can be represented either as mixtures or scale-mixtures of normals, including the Student-t and the logistic link functions (Albert and Chib 1993; Wood and Kohn 1998; Chen and Dey 2000). While sampling from the distributions in Algorithm 1 is convenient because they are of known form (i.e. uniform, truncated normal, and normal), one remaining question concerns the sampling of the cut-points

in Step 1 conditioned on the latent data. Cowles (1996) noted the possibility that in some cases, the sampling of the cutpoints conditioned on the latent data can lead to small changes in the cutpoints between successive iterations, especially as more data become available. The resulting high autocorrelation in the MCMC sample for the cutpoints could then also affect the convergence of $\boldsymbol{\beta}$. To deal with this problem, Cowles (1996) suggested that it helps to sample the latent data $\boldsymbol{z}$ and the cutpoints $\boldsymbol{\gamma}$ jointly by sampling $\boldsymbol{\gamma} \sim \pi(\boldsymbol{\gamma}|\boldsymbol{y}, \boldsymbol{\beta})$ marginalized over the latent data and subsequently sampling $\boldsymbol{z} \sim \pi(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\gamma})$, i.e. given $\boldsymbol{\gamma}$ and the remaining parameters and data. Although the resulting distribution of the cutpoints is not of standard form, Cowles employed a sequence of Metropolis-Hastings steps to sample each γ_j conditioned on $(\boldsymbol{y}, \boldsymbol{\beta}, \gamma_{j-1}, \gamma_{j+1})$. Nandram and Chen (1996) improved upon Cowles (1996) by noting that the cutpoints should be sampled jointly, not one-at-a-time, and that the particular MH proposal density suggested in Cowles (1996) may be difficult to tune. They suggested a reparameterization of the model and presented a sampler that allows for joint sampling of the reparameterized cutpoints in a single block and also marginally of the latent data using a Dirichlet proposal density that depends on the previous cutpoints, but does not depend on the other parameters or the latent data. However, Chen and Dey (2000) point out that the Dirichlet density will generally work well when the cell counts are balanced, but may fail to serve as a good proposal density when the category counts are unbalanced.

Subsequent work (e.g. Chen and Dey 2000; Albert and Chib 2001) built upon these ideas and showed that the cutpoints $\boldsymbol{\gamma}$ can easily be sampled jointly in a single block by well-tailored independence chains, marginally of $\boldsymbol{z}$, to improve the efficiency of the MCMC algorithm. Maintaining the identification restriction that $var(\varepsilon) = 1$, Albert and Chib (2001) simplified the sampling of the cutpoints $\boldsymbol{\gamma}$ by transforming them so as to remove the ordering constraint by the one-to-one map

$$\delta_j = \ln(\gamma_j - \gamma_{j-1}),\ 2 \leq j \leq J-1. \tag{6}$$

Other transformations have been considered (e.g. in Chen and Dey 2000), but details of those transformations will be delayed until Section 2.2 below as they will be related to an alternative

identification of the scale of model. In either case, however, the advantage of working with the transformed quantities $\boldsymbol{\delta} = (\delta_2, ..., \delta_{J-1})'$, instead of $\boldsymbol{\gamma} = (\gamma_2, ..., \gamma_{J-1})'$, is that the parameters of the tailored proposal density in the MH step for $\boldsymbol{\delta}$ can be obtained by unconstrained optimization and the prior $\pi(\boldsymbol{\delta})$ can be unrestricted, e.g. multivariate normal $N(\boldsymbol{\delta}_0, \boldsymbol{D}_0)$. The algorithm is defined as follows.

Algorithm 2 ***(Albert & Chib 2001) Sampling in the univariate ordered probit model (identification through variance restriction)***

1. *Sample* $\boldsymbol{\delta}, \boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\beta}$ *in one block as follows:*

   (a) *Sample* $\boldsymbol{\delta}|\boldsymbol{y}, \boldsymbol{\beta}$ *marginally of* $\boldsymbol{z}$ *by drawing* $\boldsymbol{\delta}' \sim q(\boldsymbol{\delta}|\boldsymbol{y}, \boldsymbol{\beta})$ *from a Student-t proposal density* $q(\boldsymbol{\delta}|\boldsymbol{y}, \boldsymbol{\beta}) = f_T\left(\boldsymbol{\delta}|\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{D}}, \nu\right)$, *where* $\hat{\boldsymbol{\delta}} = \arg\max f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\delta})\pi(\boldsymbol{\delta})$, $\hat{\boldsymbol{D}}$ *is the inverse of the negative Hessian of* $\ln\{f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\delta})\pi(\boldsymbol{\delta})\}$ *evaluated at* $\hat{\boldsymbol{\delta}}$, *and* ν *is a degrees of freedom parameter. Given the current value of* $\boldsymbol{\delta}$ *and the proposed draw* $\boldsymbol{\delta}'$, *return* $\boldsymbol{\delta}'$ *with probability* $\alpha_{MH}(\boldsymbol{\delta}, \boldsymbol{\delta}') = \min\left\{1, \frac{f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\delta}')\pi(\boldsymbol{\beta}, \boldsymbol{\delta}')}{f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\delta})\pi(\boldsymbol{\beta}, \boldsymbol{\delta})} \frac{f_T(\boldsymbol{\delta}|\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{D}}, \nu)}{f_T(\boldsymbol{\delta}'|\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{D}}, \nu)}\right\}$; *otherwise repeat the old value* $\boldsymbol{\delta}$.

   (b) *Sample* $\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ *by drawing* $z_i|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\gamma} \sim TN_{(\gamma_{j-1}, \gamma_j)}(\boldsymbol{x}_i'\boldsymbol{\beta}, 1)$ *for* $i = 1, ..., n$, *where* $\boldsymbol{\gamma}$ *is obtained by the one-to-one mapping used to relate* $\boldsymbol{\gamma}$ *and* $\boldsymbol{\delta}$.

2. *Sample* $\boldsymbol{\beta}|\boldsymbol{z} \sim N\left(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{B}}\right)$, *where* $\hat{\boldsymbol{B}} = (\boldsymbol{B}_0 + \boldsymbol{X}'\boldsymbol{X})^{-1}$ *and* $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{B}}\left(\boldsymbol{B}_0^{-1}\boldsymbol{\beta}_0 + \boldsymbol{X}'\boldsymbol{z}\right)$.

In Step 1 of the above algorithm, the degrees of freedom parameter ν is taken to be a low number such as 5 or 10 to ensure that the proposal density has sufficiently heavy tails. By grouping the sampling of $\boldsymbol{\delta}$ and $\boldsymbol{z}$ into a single step, the above two-block algorithm produces a well-mixing Markov chain, whose performance will be illustrated in Section 6. We next consider the issue of using alternative identification restrictions in order to fix the scale of the model. Doing so results in a different blocking of the parameters that, except for cases where there are $J = 3$ categories, will produce a three-block algorithm, as opposed to the two-block sampler given in Algorithm 2.

For this reason, in the univariate setting one can easily estimate ordinal probit models without having to consider these alternatives (except when $J = 3$), but the restrictions we discuss below are quite useful in multivariate settings. We emphasize that the choice of blocking is not unique and is something that should be determined by the researcher given the specific context of the model and data under consideration. In practice it is useful to (i) group parameters that are correlated into one block and sample them jointly, and (ii) group parameters in a way that allows for easy construction of suitable MCMC samplers.

2.2 Estimation Under Alternative Identification Restrictions

As discussed in the introduction, there are a number of ways to identify the scale of the model. To gain further insight into the identification problem, consider the ordinal probit likelihood

$$f\left(y|\beta,\gamma\right)=\prod_{i=1}^{n}\prod_{j=1}^{J}\left[\Phi\left(\frac{\gamma_j-x_i'\beta}{\sigma}\right)-\Phi\left(\frac{\gamma_{j-1}-x_i'\beta}{\sigma}\right)\right]^{1\{y_i=j\}} \tag{7}$$

The standard identification restriction commonly used in discrete data models, i.e. $var\left(\varepsilon_i\right)=\sigma^2=1$ (so that $\varepsilon_i\sim N\left(0,1\right)$), was used in Section 2.1 to preclude the possibility for arbitrary rescaling of the parameters by some constant c such that $\Phi\left(\frac{c\left(\gamma_j-\boldsymbol{x}_i'\boldsymbol{\beta}\right)}{c\sigma}\right)=\Phi\left(\frac{\gamma_j-\boldsymbol{x}_i'\boldsymbol{\beta}}{\sigma}\right)$. However, it is possible to identify the scale of the model differently. In particular, we can leave $var\left(\varepsilon_i\right)=\sigma^2$ as an unrestricted parameter to be estimated, but instead fix another cutpoint in addition to having $\gamma_0=-\infty$, $\gamma_1=0$, and $\gamma_J=\infty$ in order to determine the scale of the latent data. For instance, one possibility is to let $\gamma_2=1$. This restriction precludes the simultaneous rescaling of the numerator and denominator in $\Phi\left(\frac{\gamma_j-\boldsymbol{x}_i'\boldsymbol{\beta}}{\sigma}\right)$ because it would violate $\gamma_2=1$. In this case, using (6), one can work with $\boldsymbol{\delta}=\left(\delta_3,...,\delta_{J-1}\right)'$. Of course, any other cutpoint can be fixed instead of γ_2, and the fixing can be at any positive constant, not just 1 (e.g. Webb and Forster 2008) even though 1 is perhaps the most natural metric to use. For example, the reparameterization considered by Nandram and Chen (1996) and Chen and Dey (2000) corresponds to a particular identification scheme where $\gamma_{J-1}=1$, thus fixing the last free cutpoint instead of the second one. Under this identification restriction, the ordering constraints on the interior cutpoints can be removed by any one of a number

of transformations. Chen and Dey (2000) consider the mapping $\gamma_j = \left(\gamma_{j-1} + e^{\delta_j}\right) / \left(1 + e^{\delta_j}\right)$ which implies that

$$\delta_j = \ln \left\{ \frac{(\gamma_j - \gamma_{j-1})}{(1 - \gamma_j)} \right\}, \; 2 \leq j \leq J-2, \tag{8}$$

so that now $\boldsymbol{\delta} = (\delta_2, ..., \delta_{J-2})'$. Other transformations of $\boldsymbol{\gamma}$ to an unrestricted and real-valued $\boldsymbol{\delta}$ are conceivable and could include log-ratios of category bin-widths or trigonometric functions such as arctan and arcsin, but generally any monotone transformation from a compact set to the real line would work. Because these alternative identification schemes are isomorphic, the parameters under each identification scheme can easily be related to those under another in a one-to-one mapping. For instance, Nandram and Chen (1996) and Chen and Dey (2000) discuss how the parameters under their identification scheme relate to those under traditional identification using unit variance restrictions.

Of course, when there are only three categories ($J = 3$) and therefore only two cutpoints that separate them, these two cutpoints need not be sampled (since $\gamma_1 = 0$ and $\gamma_2 = 1$) and the different parameterizations that use fixing of a second cutpoint become identical. When $J > 3$, however, the choice of which cutpoint to fix and what mapping to apply can influence the performance of the MCMC sampler. We provide some evidence on this in Section 6. One should also note that when $J = 2$, identification can not be achieved by fixing an additional cutpoint and the model does not automatically become identical to the binary data probit model. This is because with only two outcome categories there is no γ_2 that can be fixed in order to determine the scale of the model, and without restrictions on σ^2 the model becomes unidentified. In those cases one will have to resort to the traditional identification restrictions discussed in Section 2 and Algorithm 2.

Then, for cases where $J \geq 3$ and under a semi-conjugate inverse gamma prior on σ^2, that is $\sigma^2 \sim IG\left(v_0/2, d_0/2\right)$, the resulting complete data posterior is given by

$$\begin{aligned} \pi\left(\boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{z}, \sigma^2 | \boldsymbol{y}\right) \quad \propto \quad & \left\{ \prod_{i=1}^{n} 1\left\{\gamma_{j-1} < z_i \leq \gamma_j\right\} N\left(z_i | \boldsymbol{x}_i' \boldsymbol{\beta}, \sigma^2\right) \right\} \\ & \times N\left(\boldsymbol{\beta} | \boldsymbol{\beta}_0, \boldsymbol{B}_0\right) N\left(\boldsymbol{\delta} | \boldsymbol{\delta}_0, \boldsymbol{D}_0\right) IG\left(\sigma^2 | v_0/2, d_0/2\right), \end{aligned}$$

which results in the following three-block sampling algorithm.

Algorithm 3 ***(Chen & Dey 2000) Sampling in the univariate ordered probit model (identification through cutpoint restrictions)***

1. *Sample* $\boldsymbol{\delta}, \boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\beta}$ *in one block as follows:*

   (a) *Sample* $\boldsymbol{\delta}|\boldsymbol{y}, \boldsymbol{\beta}$ *marginally of* $\boldsymbol{z}$ *by drawing* $\boldsymbol{\delta}' \sim q(\boldsymbol{\delta}|\boldsymbol{y}, \boldsymbol{\beta})$*, with* $q(\boldsymbol{\delta}|\boldsymbol{y}, \boldsymbol{\beta}) = f_T\left(\boldsymbol{\delta}|\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{D}}, \nu\right)$ *where* $\hat{\boldsymbol{\delta}} = \arg\max f\left(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\delta}, \sigma^2\right)\pi(\boldsymbol{\delta})$ *and* $\hat{\boldsymbol{D}}$ *is the inverse of the negative Hessian of* $\ln\left\{f\left(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\delta}, \sigma^2\right)\pi(\boldsymbol{\delta})\right\}$ *evaluated at* $\hat{\boldsymbol{\delta}}$*. Given the current value of* $\boldsymbol{\delta}$ *and the proposed draw* $\boldsymbol{\delta}'$*, return* $\boldsymbol{\delta}'$ *with probability* $\alpha_{MH}\left(\boldsymbol{\delta}, \boldsymbol{\delta}'\right) = \min\left\{1, \frac{f\left(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\delta}'\right)\pi(\boldsymbol{\beta}, \boldsymbol{\delta}')}{f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\delta})\pi(\boldsymbol{\beta}, \boldsymbol{\delta})} \frac{f_T\left(\boldsymbol{\delta}|\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{D}}, \nu\right)}{f_T\left(\boldsymbol{\delta}'|\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{D}}, \nu\right)}\right\}$*; otherwise repeat the old value* $\boldsymbol{\delta}$*.*

   (b) *Sample* $\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\beta}$ *by drawing* $z_i|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\delta} \sim TN_{(\gamma_{j-1}, \gamma_j)}\left(\boldsymbol{x}_i'\boldsymbol{\beta}, \sigma^2\right)$ *for* $i = 1, ..., n$*, where* $\boldsymbol{\gamma}$ *is obtained by the one-to-one mapping used to relate* $\boldsymbol{\gamma}$ *and* $\boldsymbol{\delta}$*.*

2. *Sample* $\boldsymbol{\beta}|z \sim N\left(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{B}}\right)$*, where* $\hat{\boldsymbol{B}} = \left(\boldsymbol{B}_0 + \boldsymbol{X}'\boldsymbol{X}/\sigma^2\right)^{-1}$ *and* $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{B}}\left(\boldsymbol{B}_0^{-1}\boldsymbol{\beta}_0 + \boldsymbol{X}'\boldsymbol{z}/\sigma^2\right)$*.*

3. *Sample* $\sigma^2 \sim IG\left(\frac{v_0+n}{2}, \frac{d_0+(\boldsymbol{z}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{z}-\boldsymbol{X}\boldsymbol{\beta})}{2}\right)$*.*

The above algorithm has been applied in Chen and Dey (2000), and its performance is further illustrated in Section 6. We note, however, that this is a three-block algorithm using the grouping $\left(\{\boldsymbol{\delta}, \boldsymbol{z}\}, \boldsymbol{\beta}, \sigma^2\right)$, so that generally the fitting of univariate models can be done more efficiently using the two-block sampler in Algorithm 2 that was developed under standard identification restrictions and used the blocking $(\{\boldsymbol{\delta}, \boldsymbol{z}\}, \boldsymbol{\beta})$. (We note that even when $J = 3$ and $\boldsymbol{\delta}$ need not be sampled, Algorithm 3 is still a three-block algorithm involving $\left(\boldsymbol{z}, \boldsymbol{\beta}, \sigma^2\right)$, but it does not involve an MH step for $\boldsymbol{\delta}$.) Algorithm 2 is also more useful when it comes to model comparison as it allows for an easier computation of the marginal likelihood, which will be discussed in Section 4. Nonetheless, the ideas behind Algorithm 3 can be quite useful when applied to the multivariate setting as discussed in Section 3 below.

In closing, we mention that another approach can be used to identify the scale of the model. We mention this approach only for the sake of completeness and not because we endorse it, as it imposes additional restrictions not required by the other identification approaches discussed so far. In particular, in (7) we first identified the model by fixing $\sigma^2 = 1$, leading to the standard version of the ordered probit model. We subsequently noted that arbitrary rescaling can be prevented by fixing an additional cutpoint and suggested that fixing $\gamma_2 = 1$ or $\gamma_{J-1} = 1$, in addition to the usual $\gamma_1 = 0$, could be implemented. But by considering (7), one can easily see that the potential for arbitrary rescaling can be removed by fixing one of the elements of $\boldsymbol{\beta}$, say $\beta_h = 1$. While this formally identifies the likelihood without the need for restricting $\gamma_2, ..., \gamma_{J-1}$ or σ^2, this identification restriction imposes both a scale restriction and a sign restriction because in reality even if $\beta_h \neq 0$, we might mistakenly assign a positive effect on β_h by fixing it at 1, when its effect could be negative, so that $\beta_h = -1$ would have been appropriate). Moreover, this restriction complicates the analysis when one is interested in performing model comparison tests whereby x_h may be removed from the set of covariates (so that β_h is no longer part of the model), requiring that normalization be based on a different covariate effect.

3 Multivariate Ordinal Outcomes

We now extend the preceding discussion to the case of multivariate ordinal outcomes. The framework for this analysis follows closely that of Chib and Greenberg (1998) and Chen and Dey (2000), who dealt with multivariate binary and ordinal outcomes, respectively. To introduce the setting, we write the multivariate version of the ordinal probit model using a threshold-crossing representation for the latent variables

$$\boldsymbol{z}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i,$$

where the q-dimensional vector of latent variables $\boldsymbol{z}_i = (z_{i1}, ..., z_{iq})'$ implies a vector of observed responses $\boldsymbol{y}_i$ according to the discretization imposed by the variable-specific cutpoints, namely

$$y_{ik} = j \quad \text{if} \quad \gamma_{k,j-1} < z_{ik} \leq \gamma_{kj}, \quad \text{for} \quad i = 1, ..., n, \quad k = 1, ..., q. \tag{9}$$

In the multivariate version of the ordered probit model, the matrix of covariates

$$\boldsymbol{X}_i = \begin{pmatrix} \begin{pmatrix} x_{i11} & \cdots & x_{i1k_1} \end{pmatrix} & & & \\ & \begin{pmatrix} x_{i21} & \cdots & x_{i2k_2} \end{pmatrix} & & \\ & & \ddots & \\ & & & \begin{pmatrix} x_{iq1} & \cdots & x_{iqk_q} \end{pmatrix} \end{pmatrix}$$

is in the form used in seemingly unrelated regression models, $\boldsymbol{\beta}$ is a k-dimensional vector of corresponding regression parameters (with $k = \sum_{j=1}^{q} k_j$), and $\boldsymbol{\varepsilon}_i$ is a q-dimensional vector of disturbances where $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Omega})$ with $\boldsymbol{\Omega}$ being a symmetric positive definite covariance matrix. It should be noted that in the model, each element of $\boldsymbol{y}_i$ can have a different number of categories $\boldsymbol{J} = (J_1, ..., J_q)'$ and its own set of cutpoints $\boldsymbol{\gamma}_k = (\gamma_1, ..., \gamma_{J_k})'$ for $k = 1, ..., q$, that can be collected in $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', ..., \boldsymbol{\gamma}_q')'$. Also, because $\boldsymbol{\Omega}$ is not required to be diagonal, the model can account for the presence of unobserved factors that can correlate the latent variables in the vector $\boldsymbol{z}_i$. Despite these correlations, however, sampling of the multivariate model can be done as a straightforward extension of the univariate algorithms by using the fact that the conditional distributions of the multivariate Gaussian distribution are also Gaussian with well-known moments. Moreover, the probit model allows for further modeling extensions – e.g. Chen and Dey (2000) extend the basic setup to other link functions such as the class of scale mixtures of multivariate normal densities that is of interest in itself or can be used to approximate other link functions such as the multivariate t-link and logit models (see also O'Brien and Dunson 2004, who develop a multivariate logit model for ordinal outcomes). For this reason, the multivariate probit model appears to be a desirable model that strikes a balance between flexibility and conceptual simplicity.

To identify the model, we use parallel restrictions to those discussed in Section 2, imposed on each of the marginal distributions for the individual z_{ik}, $k = 1, ..., q$. In particular, following the layout of Section 2, we first present an algorithm in which the scale of the individual (marginal) latent processes is fixed by imposing a unit error variance, leading to a matrix $\boldsymbol{\Omega}$ that is in correlation form as in the multivariate probit model for binary data in Chib and Greenberg (1998). Subsequently, we discuss an alternative algorithm, where each of the marginal distributions is scaled

by fixing two cutpoints (e.g. either $\gamma_{k1} = 0$ and $\gamma_{k2} = 1$, or $\gamma_{k1} = 0$ and $\gamma_{k,J-1} = 1$, $k = 1, ..., q$), so that the covariance matrix $\boldsymbol{\Omega}$ is free and can be sampled from a known full-conditional distribution.

When $\boldsymbol{\Omega}$ in correlation form, the sampling of the correlations is non-standard and we approach this task by relying on the versatility of the ARMH algorithm. To introduce that algorithm, let $\boldsymbol{\theta}$ be a parameter vector whose density, $\pi(\boldsymbol{\theta})$, is the target density of interest, but is possibly known only up to a normalizing constant and is not easy to simulate. Let $h(\boldsymbol{\theta})$ denote a source (or proposal) density for the ARMH algorithm and let the constant c define the region of domination

$$\mathcal{D} = \{\boldsymbol{\theta} : \pi(\boldsymbol{\theta}) \leq ch(\boldsymbol{\theta})\}$$

which is a subset of the support Θ of the target density (because the domination condition need not be satisfied for all $\boldsymbol{\theta} \in \Theta$, the source density $h(\boldsymbol{\theta})$ is often called a pseudo-dominating density). Commonly, the choice of a pseudo-dominating density is determined by tailoring. For instance, $h(\boldsymbol{\theta})$ can be given by a multivariate-t density $h(\boldsymbol{\theta}) = f_T(\boldsymbol{\theta}|\boldsymbol{\mu}, \tau\boldsymbol{V}, \nu)$ with ν degrees of freedom, mean $\boldsymbol{\mu}$ given by the maximum of the target density $\pi(\boldsymbol{\theta})$, and scale matrix $\tau\boldsymbol{V}$, where $\boldsymbol{V}$ is the inverse of the negative Hessian of $\ln \pi(\boldsymbol{\theta})$ evaluated at $\boldsymbol{\mu}$, and τ is a parameter allowing for the possibility of additional tuning. Let $\mathcal{D}^c$ be the complement of $\mathcal{D}$, and suppose that the current state of the chain is $\boldsymbol{\theta}$. Then the ARMH algorithm proceeds as follows.

Algorithm 4 ***The accept-reject Metropolis-Hastings (ARMH) algorithm***

1. *A-R step: Generate a draw $\boldsymbol{\theta}' \sim h(\boldsymbol{\theta})$; accept $\boldsymbol{\theta}'$ with probability $\alpha_{AR}(\boldsymbol{\theta}') = \min\left\{1, \frac{\pi(\boldsymbol{\theta}')}{ch(\boldsymbol{\theta}'|\boldsymbol{y})}\right\}$. Continue the process until a draw $\boldsymbol{\theta}'$ has been accepted.*

2. *M-H step: Given the current value $\boldsymbol{\theta}$ and the proposed value $\boldsymbol{\theta}'$:*

   (a) if $\boldsymbol{\theta} \in \mathcal{D}$, set $\alpha_{MH}(\boldsymbol{\theta}, \boldsymbol{\theta}') = 1$;

   (b) if $\boldsymbol{\theta} \in \mathcal{D}^c$ and $\boldsymbol{\theta}' \in \mathcal{D}$, set $\alpha_{MH}(\boldsymbol{\theta}, \boldsymbol{\theta}') = \frac{ch(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta})}$;

   (c) if $\boldsymbol{\theta} \in \mathcal{D}^c$ and $\boldsymbol{\theta}' \in \mathcal{D}^c$, set $\alpha_{MH}(\boldsymbol{\theta}, \boldsymbol{\theta}') = \min\left\{1, \frac{\pi(\boldsymbol{\theta}')h(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta})h(\boldsymbol{\theta}')}\right\}$.

Return $\boldsymbol{\theta}'$ with probability $\alpha_{MH}(\boldsymbol{\theta}, \boldsymbol{\theta}')$; otherwise return $\boldsymbol{\theta}$.

The ARMH algorithm is an MCMC sampling procedure which nests both the accept-reject and MH algorithms when $\mathcal{D}^c$ and $\mathcal{D}$ become empty sets, respectively. But in the intermediate case when both $\mathcal{D}$ and $\mathcal{D}^c$ are non-empty, ARMH has several attractive features that make it a useful choice for our setting. First, the algorithm is well suited to problems that do not require conjugacy and result in non-standard full-conditional densities, which is the case for the elements of $\boldsymbol{\Omega}$ when it is in correlation form. Second, ARMH can be less demanding and works quite well even if the proposal density $h(\boldsymbol{\theta})$ is only a rough approximation of the target density (e.g. Chib and Jeliazkov 2005). This is particularly useful in our setting because previous research suggests that standard asymptotic approximating densities can be only rough approximations when sample sizes are small (e.g. Zellner and Rossi 1984). Third, ARMH can produce draws that are closer to iid than those from a similarly constructed MH simulator, but without requiring global domination that is needed for the simple accept-reject algorithm. Fourth, in sampling covariance or correlation matrices, only draws that satisfy positive definiteness pass through the A-R step of the algorithm, thus improving the performance of the MH step. Finally, the building blocks of the ARMH algorithm provide a straightforward way to estimate the marginal likelihood (e.g. Chib and Jeliazkov 2005), which will be discussed in Section 4.

We are now ready to proceed with estimation of the multivariate ordered probit under traditional identification restrictions, where the scale is fixed by requiring $\boldsymbol{\Omega}$ to be in correlation form. We begin by considering the complete data posterior $\pi(\boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\rho}, \boldsymbol{z}|\boldsymbol{y})$, where $\boldsymbol{\rho}$ is the vector of unique correlations in $\boldsymbol{\Omega}$. Assuming the prior $\boldsymbol{\rho} \sim N(\boldsymbol{\rho}_0, \boldsymbol{R}_0) 1\{\boldsymbol{\rho} \in S\}$, where S is the set of correlations that produce a positive definite matrix $\boldsymbol{\Omega}$ with ones on the main diagonal, we have

$$\begin{aligned}\pi(\boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\rho}, \boldsymbol{z}|\boldsymbol{y}) \quad \propto \quad & \left\{\prod_{i=1}^{n}\left[\prod_{k=1}^{q} 1\{\gamma_{k,j-1} < z_{ik} \leq \gamma_{kj}\}\right] N(\boldsymbol{z}_i|\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Omega})\right\} \\ & \times N(\boldsymbol{\beta}|\boldsymbol{\beta}_0, \boldsymbol{B}_0) N(\boldsymbol{\delta}|\boldsymbol{\delta}_0, \boldsymbol{D}_0) N(\boldsymbol{\rho}|\boldsymbol{r}_0, \boldsymbol{R}_0) 1\{\boldsymbol{\rho} \in S\},\end{aligned}$$

where the index j in the indicator functions above is determined by the value of y_{ik} according

to (9). The above posterior distribution gives rise to the following MCMC estimation algorithm, where as a matter of notation, we will use "$\backslash k$" to represent all elements in a set except the kth one (e.g. if $\boldsymbol{\theta}_k$ is the kth block in the vector $\boldsymbol{\theta} = \left(\boldsymbol{\theta}_1', ..., \boldsymbol{\theta}_p'\right)'$, then $\boldsymbol{\theta}_{\backslash k} = \left(\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_{k-1}', \boldsymbol{\theta}_{k+1}', \ldots, \boldsymbol{\theta}_p\right)'$).

Algorithm 5 ***Sampling in the multivariate ordered probit model ($\boldsymbol{\Omega}$ is in correlation form)***

1. *For $k = 1, ..., q$, sample $\boldsymbol{\delta}_k, \boldsymbol{z}_k | \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{z}_{\backslash k}$ in one block as follows:*

   (a) *Sample $\boldsymbol{\delta}_k | \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{z}_{\backslash k}$ marginally of $\boldsymbol{z}_k$ by drawing a value $\boldsymbol{\delta}_k' \sim q\left(\boldsymbol{\delta}_k | \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{z}_{\backslash k}\right)$, where $q\left(\boldsymbol{\delta}_k | \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{z}_{\backslash k}\right) = f_T\left(\boldsymbol{\delta}_k | \hat{\boldsymbol{\delta}}_k, \hat{\boldsymbol{D}}_k, \nu\right)$ with $\hat{\boldsymbol{\delta}}_k = \arg\max f\left(\boldsymbol{y}_k | \boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\delta}, \boldsymbol{z}_{\backslash k}\right) \pi\left(\boldsymbol{\delta}_k | \boldsymbol{\delta}_{\backslash k}\right)$ and $\hat{\boldsymbol{D}}_k$ is the inverse of the negative Hessian of $\ln\left\{f\left(\boldsymbol{y}_k | \boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\delta}, \boldsymbol{z}_{\backslash k}\right) \pi\left(\boldsymbol{\delta}_k | \boldsymbol{\delta}_{\backslash k}\right)\right\}$ evaluated at $\hat{\boldsymbol{\delta}}_k$. Given the current value of $\boldsymbol{\delta}$ and the proposed draw $\boldsymbol{\delta}'$, return $\boldsymbol{\delta}'$ with probability*

$$\alpha_{MH}\left(\boldsymbol{\delta}_k, \boldsymbol{\delta}_k'\right) = \min\left\{1, \frac{f\left(\boldsymbol{y}_k | \boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\delta}_k', \boldsymbol{z}_{\backslash k}\right) \pi\left(\boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\delta}_k' | \boldsymbol{\delta}_{\backslash k}\right)}{f\left(\boldsymbol{y}_k | \boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\delta}, \boldsymbol{z}_{\backslash k}\right) \pi\left(\boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\delta}_k | \boldsymbol{\delta}_{\backslash k}\right)} \frac{f_T(\boldsymbol{\delta}_k | \hat{\boldsymbol{\delta}}_k, \hat{\boldsymbol{D}}, \nu)}{f_T(\boldsymbol{\delta}_k' | \hat{\boldsymbol{\delta}}_k, \hat{\boldsymbol{D}}, \nu)}\right\},$$

   and otherwise repeat the old value $\boldsymbol{\delta}$.

   (b) *Sample $\boldsymbol{z}_k | \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\delta}_k, \boldsymbol{z}_{\backslash k}$ by drawing $z_{ik} | \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{\gamma}, \boldsymbol{z}_{\backslash k} \sim TN_{(\gamma_{j-1}, \gamma_j)}\left(\mu_{k|\backslash k}, \sigma^2_{k|\backslash k}\right)$ for $i = 1, ..., n$, where $\mu_{k|\backslash k}$ and $\sigma^2_{k|\backslash k}$ are the usual conditional mean and variance for a Gaussian distribution.*

2. *Sample $\boldsymbol{\beta} | \boldsymbol{z} \sim N\left(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{B}}\right)$, where $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{B}}$ are given by $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{B}}\left(\boldsymbol{B}_0^{-1}\boldsymbol{\beta}_0 + \sum_{i=1}^n \boldsymbol{X}_i'\boldsymbol{\Omega}^{-1}\boldsymbol{z}_i\right)$ and $\hat{\boldsymbol{B}} = \left(\boldsymbol{B}_0^{-1} + \sum_{i=1}^n \boldsymbol{X}_i'\boldsymbol{\Omega}^{-1}\boldsymbol{X}_i\right)^{-1}$.*

3. *Sample $\boldsymbol{\rho} | \boldsymbol{z}, \boldsymbol{\beta}$ by ARMH (Algorithm 4) with proposal density $h\left(\boldsymbol{\rho} | \boldsymbol{z}, \boldsymbol{\beta}\right) = f_T\left(\boldsymbol{\rho} | \hat{\boldsymbol{\rho}}, \tau\hat{\boldsymbol{R}}, \nu\right)$, where $\hat{\boldsymbol{\rho}}$ and $\hat{\boldsymbol{R}}$ are approximations to the maximizer and inverse of the negative Hessian of $\ln\left\{f\left(z | \beta, \rho\right) \pi\left(\rho\right)\right\} = \ln\left\{\prod_{i=1}^n N\left(\boldsymbol{z}_i | \boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Omega}\right) N\left(\boldsymbol{\rho} | \boldsymbol{\rho}_0, \boldsymbol{R}_0\right) 1\left\{\boldsymbol{\rho} \in S\right\}\right\}$, respectively, and τ is a tuning constant:*

(a) *as a first step, try* $\hat{\boldsymbol{\rho}} = \hat{\boldsymbol{R}}\left(\boldsymbol{R}_0^{-1}\boldsymbol{\rho}_0 + \boldsymbol{C}^{-1}\boldsymbol{c}\right)$ *and* $\hat{\boldsymbol{R}} = \left(\boldsymbol{R}_0^{-1} + \boldsymbol{C}^{-1}\right)^{-1}$ *where* $\boldsymbol{c}$ *is the vector of unique elements of corr* $(\boldsymbol{z}_i - \boldsymbol{X}_i\boldsymbol{\beta})$ *and* $\boldsymbol{C}$ *is the BHHH outer product of gradients estimate of the modal dispersion matrix evaluated at* $\boldsymbol{c}$*;*

(b) *if the gradient at* $\hat{\boldsymbol{\rho}}$ *above is far from zero (so that* $\hat{\boldsymbol{\rho}}$ *is not a good approximation), fine-tune* $\hat{\boldsymbol{\rho}}$ *by further optimization and proceed with the ARMH algorithm.*

Computing the quantities in Step 3a of Algorithm 5 is particularly easy and does not require optimization; moreover, it is very fast because the sampling is conditional on $\boldsymbol{z}$ so that the computations are comparable to those in a continuous data model. Importantly, when fine-tuning is required, it can generally be accomplished in a few steps by quasi-Newton methods. Alternatively, the proposal density in the ARMH step can be initialized with $\hat{\boldsymbol{\rho}}$ being the correlations from the last draw of $\boldsymbol{\Omega}$ and $\hat{\boldsymbol{R}}$ being the BHHH estimate of modal dispersion matrix at $\hat{\boldsymbol{\rho}}$.

Because of the standard way of identifying the parameters, the multivariate ordinal probit model simplifies to the multivariate probit model for binary data that was analyzed in Chib and Greenberg (1998) when there are only two choice categories. Moreover, with this algorithm it is possible to fit models with restricted covariance matrices that may involve off-diagonal zeros or various Toeplitz structures in $\boldsymbol{\Omega}$. The performance of this algorithm is demonstrated in Section 6.

We now turn attention to the second identification scheme that relies on fixing two cutpoints for each outcome such as $\gamma_{k1} = 0$ and either $\gamma_{k2} = 1$ or $\gamma_{k,J-1} = 1$ for $k = 1, ..., q$. As discussed in the univariate case, doing so frees up the variances of the latent variables, thus removing the requirement that the diagonal elements of $\boldsymbol{\Omega}$ must be ones. The benefit of doing so is that when $\boldsymbol{\Omega}$ is unrestricted, we can use the usual semi-conjugate inverse Wishart prior on $\boldsymbol{\Omega}$, that is $\boldsymbol{\Omega} \sim IW\left(v_0, \boldsymbol{W}_0\right)$, which produces the a complete data posterior that is given by

$$\begin{aligned}\pi\left(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{z}|\boldsymbol{y}\right) \quad \propto \quad & \left\{\prod_{i=1}^{n}\left[\prod_{k=1}^{q} 1\left\{\gamma_{k,j-1} < z_{ik} \leq \gamma_{kj}\right\}\right] N\left(\boldsymbol{z}_i|\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{\Omega}\right)\right\} \\ & \times N\left(\boldsymbol{\beta}|\boldsymbol{\beta}_0, \boldsymbol{B}_0\right) N\left(\boldsymbol{\delta}|\boldsymbol{\delta}_0, \boldsymbol{D}_0\right) IW\left(\boldsymbol{\Omega}|v_0, \boldsymbol{W}_0\right).\end{aligned}$$

This posterior distribution gives rise to the following MCMC estimation algorithm (note that

because a second cutpoint is fixed for each response variable, the vectors $\boldsymbol{\delta}_k$ for $k = 1, ..., q$, have one less element than the corresponding vectors in Algorithm 5).

Algorithm 6 ***(Chen & Dey 2000) Sampling in the multivariate ordered probit model (identification through cutpoint restrictions)***

1. *For $k = 1, ..., q$, sample $\boldsymbol{\delta}_k, \boldsymbol{z}_k | \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{z}_{\backslash k}$ in one block as follows*

   (a) *Sample $\boldsymbol{\delta}_k | \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\Omega}, \boldsymbol{z}_{\backslash k}$ marginally of $\boldsymbol{z}_k$ by drawing $\boldsymbol{\delta}_k' \sim q\left(\boldsymbol{\delta}_k | \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\Omega}, \boldsymbol{z}_{\backslash k}\right)$, where $q\left(\boldsymbol{\delta}_k | \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\Omega}, \boldsymbol{z}_{\backslash k}\right) = f_T\left(\boldsymbol{\delta}_k | \hat{\boldsymbol{\delta}}_k, \hat{\boldsymbol{D}}_k, \nu\right)$ with $\hat{\boldsymbol{\delta}}_k = \arg\max f\left(\boldsymbol{y}_k | \boldsymbol{\beta}, \boldsymbol{\Omega}, \boldsymbol{\delta}, \boldsymbol{z}_{\backslash k}\right) \pi\left(\boldsymbol{\delta}_k | \boldsymbol{\delta}_{\backslash k}\right)$ and $\hat{\boldsymbol{D}}_k$ is the inverse of the negative Hessian of $\ln\left\{f\left(\boldsymbol{y}_k | \boldsymbol{\beta}, \boldsymbol{\Omega}, \boldsymbol{\delta}, \boldsymbol{z}_{\backslash k}\right) \pi\left(\boldsymbol{\delta}_k | \boldsymbol{\delta}_{\backslash k}\right)\right\}$ evaluated at $\hat{\boldsymbol{\delta}}_k$. Given the current value $\boldsymbol{\delta}$ and the proposed draw $\boldsymbol{\delta}'$, return $\boldsymbol{\delta}'$ with probability*

$$\alpha_{MH}\left(\boldsymbol{\delta}_k, \boldsymbol{\delta}_k'\right) = \min\left\{1, \frac{f\left(\boldsymbol{y}_k | \boldsymbol{\beta}, \boldsymbol{\Omega}, \boldsymbol{\delta}_k', \boldsymbol{z}_{\backslash k}\right) \pi\left(\boldsymbol{\beta}, \boldsymbol{\Omega}, \boldsymbol{\delta}_k' | \boldsymbol{\delta}_{\backslash k}\right)}{f\left(\boldsymbol{y}_k | \boldsymbol{\beta}, \boldsymbol{\Omega}, \boldsymbol{\delta}, \boldsymbol{z}_{\backslash k}\right) \pi\left(\boldsymbol{\beta}, \boldsymbol{\Omega}, \boldsymbol{\delta}_k | \boldsymbol{\delta}_{\backslash k}\right)} \frac{f_T(\boldsymbol{\delta}_k | \hat{\boldsymbol{\delta}}_k, \hat{\boldsymbol{D}}, \nu)}{f_T(\boldsymbol{\delta}_k' | \hat{\boldsymbol{\delta}}_k, \hat{\boldsymbol{D}}, \nu)}\right\},$$

   and otherwise repeat the old value $\boldsymbol{\delta}$.

   (b) *Sample $\boldsymbol{z}_k | \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\Omega}, \boldsymbol{\delta}_k, \boldsymbol{z}_{\backslash k}$ by drawing $z_{ik} | \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\Omega}, \boldsymbol{\gamma}, \boldsymbol{z}_{\backslash k} \sim TN_{(\gamma_{j-1}, \gamma_j)}\left(\mu_{k|\backslash k}, \sigma^2_{k|\backslash k}\right)$ for $i = 1, ..., n$, where $\mu_{k|\backslash k}$ and $\sigma^2_{k|\backslash k}$ are the usual conditional mean and variance for a Gaussian distribution.*

2. *Sample $\boldsymbol{\beta} | \boldsymbol{z} \sim N\left(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{B}}\right)$, where $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{B}}$ are given by $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{B}}\left(\boldsymbol{B}_0^{-1}\boldsymbol{\beta}_0 + \sum_{i=1}^n \boldsymbol{X}_i'\boldsymbol{\Omega}^{-1}\boldsymbol{z}_i\right)$ and $\hat{\boldsymbol{B}} = \left(\boldsymbol{B}_0^{-1} + \sum_{i=1}^n \boldsymbol{X}_i'\boldsymbol{\Omega}^{-1}\boldsymbol{X}_i\right)^{-1}$.*

3. *Sample $\boldsymbol{\Omega} \sim IW\left(v_0 + n, \boldsymbol{W}_0 + (\boldsymbol{z} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{z} - \boldsymbol{X}\boldsymbol{\beta})\right)$.*

The above algorithm makes the conditional sampling of $\boldsymbol{\Omega}$ particularly attractive since the draws are obtained from a well-known full-conditional distribution, so that sampling is fast and easy to implement. Another benefit of this algorithm is that no cutpoints need to be sampled for any response variable that has three categories. Unfortunately, the identification restriction can not be applied (a) when one or more elements in $\boldsymbol{y}_i$ are binary and there is only one cutpoint

between the two categories so that unit variance is required for identification, (b) when $\boldsymbol{\Omega}$ must be restricted in any other way (e.g. when a block of off-diagonal elements is restricted to zero), (c) when different subsets of elements of $\boldsymbol{\Omega}$ are updated by different subsamples of data (e.g. in the case of missing or incidentally truncated outcomes), or (d) when non-conjugate priors are used to model the elements of the matrix $\boldsymbol{\Omega}$. In these cases, estimation can proceed under traditional identification assumptions with $\boldsymbol{\Omega}$ being a correlation matrix.

4 Model Comparison

A central issue in the analysis of statistical data is model formulation, since the appropriate specification is rarely known and is subject to uncertainty. Among other considerations, the uncertainty may be due to the problem of variable selection (i.e. the specific covariates to be included in the model) or perhaps due to the functional specification through which the covariates affect the probability of the outcome. In general, given the data $\boldsymbol{y}$, interest centers upon a collection of models $\{\mathcal{M}_1, ..., \mathcal{M}_L\}$ representing competing hypotheses about $\boldsymbol{y}$. Each model $\mathcal{M}_l$ is characterized by a model-specific parameter vector $\boldsymbol{\theta}_l$ and sampling density $f(\boldsymbol{y}|\mathcal{M}_l, \boldsymbol{\theta}_l)$. Bayesian model selection proceeds by comparison of the models in $\{\mathcal{M}_l\}$ through their posterior odds ratio, which for any two models $\mathcal{M}_i$ and $\mathcal{M}_j$ is written as

$$\frac{\Pr(\mathcal{M}_i|\boldsymbol{y})}{\Pr(\mathcal{M}_j|\boldsymbol{y})} = \frac{\Pr(\mathcal{M}_i)}{\Pr(\mathcal{M}_j)} \times \frac{m(\boldsymbol{y}|\mathcal{M}_i)}{m(\boldsymbol{y}|\mathcal{M}_j)}\,, \tag{10}$$

where $m(\boldsymbol{y}|\mathcal{M}_l) = \int f(\boldsymbol{y}|\mathcal{M}_l, \boldsymbol{\theta}_l)\pi_l(\boldsymbol{\theta}_l|\mathcal{M}_l)d\boldsymbol{\theta}_l$ is the marginal likelihood of $\mathcal{M}_l$. The first fraction on the right hand side of (10) is known as the prior odds and the second as the Bayes factor. We show that the question of calculating the marginal likelihood for ordinal data models under each of the identification schemes discussed above can be managed through straightforward application of methods that are built upon the structure of the sampling algorithms.

Chib (1995) provides a method based on the recognition that the marginal likelihood can be

re-expressed as

$$m(\boldsymbol{y}|\mathcal{M}_l) = \frac{f(\boldsymbol{y}|\mathcal{M}_l, \boldsymbol{\theta}_l)\pi(\boldsymbol{\theta}_l|\mathcal{M}_l)}{\pi(\boldsymbol{\theta}_l|\boldsymbol{y}, \mathcal{M}_l)}, \tag{11}$$

which holds for any point $\boldsymbol{\theta}_l$, so that calculation of the marginal likelihood is reduced to finding an estimate of the posterior ordinate $\pi(\boldsymbol{\theta}_l^*|\boldsymbol{y}, \mathcal{M}_l)$ at a single point $\boldsymbol{\theta}_l^*$, given that usually $f(\boldsymbol{y}|\mathcal{M}_l, \boldsymbol{\theta}_l^*)$ and $\pi(\boldsymbol{\theta}_l^*|\mathcal{M}_l)$ for the numerator of (11) are either available directly or by an alternative calculation. In practice, the point $\boldsymbol{\theta}_l^*$ is often taken as the posterior mean or mode, which tends to minimize the estimation variability. In the following, we suppress the model index for notational convenience.

In the context of ordinal probit models, the parameter vector $\boldsymbol{\theta}$ will consist of the regression coefficients $\boldsymbol{\beta}$, the cutpoint transformations $\boldsymbol{\delta}$, and possibly σ^2 or $\boldsymbol{\Omega}$ (depending on the identification scheme and the dimensionality of $\boldsymbol{y}_i$). To keep the discussion general, let $\boldsymbol{\theta}$ be split into B components or blocks that emerge in constructing the MCMC sampler under the chosen identification scheme, so that $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_B)$. Let $\boldsymbol{\psi}_i^* = (\boldsymbol{\theta}_1^*, ..., \boldsymbol{\theta}_i^*)$ denote the blocks up to i, fixed at their values in $\boldsymbol{\theta}^*$, and let $\boldsymbol{\psi}^{i+1} = (\boldsymbol{\theta}_{i+1}, ..., \boldsymbol{\theta}_B)$ denote the blocks beyond i. Then, by the law of total probability we have

$$\pi\left(\boldsymbol{\theta}_1^*, ..., \boldsymbol{\theta}_B^*|\boldsymbol{y}\right) = \prod_{i=1}^{B} \pi\left(\boldsymbol{\theta}_i^*|\boldsymbol{y}, \boldsymbol{\theta}_1^*, ..., \boldsymbol{\theta}_{i-1}^*\right) = \prod_{i=1}^{B} \pi\left(\boldsymbol{\theta}_i^*|\boldsymbol{y}, \boldsymbol{\psi}_{i-1}^*\right) .$$

When the full-conditional densities are known, each ordinate $\pi(\boldsymbol{\theta}_i^*|\boldsymbol{y}, \boldsymbol{\psi}_{i-1}^*)$ can be estimated by Rao-Blackwellization as $\pi\left(\boldsymbol{\theta}_i^*|\boldsymbol{y}, \boldsymbol{\psi}_{i-1}^*\right) \approx J^{-1}\sum_{j=1}^{J} \pi\left(\boldsymbol{\theta}_i^*|\boldsymbol{y}, \boldsymbol{\psi}_{i-1}^*, \boldsymbol{\psi}^{i+1,(j)}\right)$, where $\boldsymbol{\psi}^{i,(j)} \sim \pi\left(\boldsymbol{\psi}^i|\boldsymbol{y}, \boldsymbol{\psi}_{i-1}^*\right)$, $j = 1, ..., J$, come from a *reduced run*, where sampling is only over $\boldsymbol{\psi}^i$, with the blocks $\boldsymbol{\psi}_{i-1}^*$ being held fixed. The ordinate $\pi\left(\boldsymbol{\theta}_1^*|y\right)$ for the first block of parameters $\boldsymbol{\theta}_1$ is estimated with draws $\boldsymbol{\theta} \sim \pi\left(\boldsymbol{\theta}|\boldsymbol{y}\right)$ from the main MCMC run, while the ordinate $\pi\left(\boldsymbol{\theta}_B^*|\boldsymbol{y}, \boldsymbol{\psi}_{B-1}^*\right)$ is available directly.

When one or more of the full conditional densities are not of a standard form and sampling requires the MH algorithm, Chib and Jeliazkov (2001) use the local reversibility of the MH chain to show that

$$\pi(\boldsymbol{\theta}_i^*|\boldsymbol{y}, \boldsymbol{\psi}_{i-1}^*) = \frac{E_1\left\{\alpha_{MH}\left(\boldsymbol{\theta}_i, \boldsymbol{\theta}_i^*|\boldsymbol{y}, \boldsymbol{\psi}_{i-1}^*, \boldsymbol{\psi}^{i+1}\right) q\left(\boldsymbol{\theta}_i, \boldsymbol{\theta}_i^*|\boldsymbol{y}, \boldsymbol{\psi}_{i-1}^*, \boldsymbol{\psi}^{i+1}\right)\right\}}{E_2\left\{\alpha_{MH}\left(\boldsymbol{\theta}_i^*, \boldsymbol{\theta}_i|\boldsymbol{y}, \boldsymbol{\psi}_{i-1}^*, \boldsymbol{\psi}^{i+1}\right)\right\}}, \tag{12}$$

where E_1 is the expectation with respect to conditional posterior $\pi(\boldsymbol{\psi}^i|\boldsymbol{y},\boldsymbol{\psi}^*_{i-1})$ and E_2 that with respect to $\pi(\boldsymbol{\psi}^{i+1}|y,\boldsymbol{\psi}^*_i)q(\boldsymbol{\theta}^*_i,\ \boldsymbol{\theta}_i|\boldsymbol{y},\boldsymbol{\psi}^*_{i-1},\boldsymbol{\psi}^{i+1})$. In the preceding, $q(\boldsymbol{\theta},\boldsymbol{\theta}'|y)$ denotes the candidate generating density of the MH chain for moving from the current value $\boldsymbol{\theta}$ to a proposed value $\boldsymbol{\theta}'$, and $\alpha_{MH}(\boldsymbol{\theta}_i,\boldsymbol{\theta}'_i|\boldsymbol{y},\boldsymbol{\psi}^*_{i-1},\boldsymbol{\psi}^{i+1})$ denotes the MH probability of move from $\boldsymbol{\theta}$ to $\boldsymbol{\theta}'$. For blocks that are sampled using the multi-block version of the ARMH algorithm, Chib and Jeliazkov (2005) show that

$$\pi(\boldsymbol{\theta}^*_i|\boldsymbol{y},\boldsymbol{\psi}^*_{i-1})=\frac{E_1\left\{\alpha_{MH}\left(\boldsymbol{\theta}_i,\boldsymbol{\theta}^*_i|\boldsymbol{y},\boldsymbol{\psi}^*_{i-1},\boldsymbol{\psi}^{i+1}\right)\alpha_{AR}\left(\boldsymbol{\theta}^*_i|\boldsymbol{y},\boldsymbol{\psi}^*_{i-1},\boldsymbol{\psi}^{i+1}\right)h\left(\boldsymbol{\theta}^*_i|\boldsymbol{y},\boldsymbol{\psi}^*_{i-1},\boldsymbol{\psi}^{i+1}\right)\right\}}{E_2\left\{\alpha_{MH}\left(\boldsymbol{\theta}^*_i,\boldsymbol{\theta}_i|\boldsymbol{y},\boldsymbol{\psi}^*_{i-1},\boldsymbol{\psi}^{i+1}\right)\alpha_{AR}\left(\boldsymbol{\theta}_i|\boldsymbol{y},\boldsymbol{\psi}^*_{i-1},\boldsymbol{\psi}^{i+1}\right)\right\}}, \tag{13}$$

where, similarly to above, E_1 is the expectation with respect to $\pi(\boldsymbol{\psi}^i|\boldsymbol{y},\boldsymbol{\psi}^*_{i-1})$ and E_2 that with respect to $\pi(\boldsymbol{\psi}^{i+1}|y,\boldsymbol{\psi}^*_i)h(\ \boldsymbol{\theta}_i|\boldsymbol{y},\boldsymbol{\psi}^*_{i-1},\boldsymbol{\psi}^{i+1})$ with $h(\cdot)$ being the proposal density in the AR step of the ARMH algorithm. Each of these expectations can be computed from the output of appropriately chosen reduced runs, where $\boldsymbol{\psi}^*_{i-1}$ is kept fixed. Methods for evaluating the variability of the ordinate estimates in each of the above cases are presented in Chib (1995) and Chib and Jeliazkov (2001, 2005).

Estimation of the marginal likelihoods for univariate ordinal data models is quite straightforward, regardless of which identification scheme is used. In particular, when the model is identified by assuming $var(\varepsilon)=1$, then the marginal likelihood can be estimated by using the posterior decomposition

$$\pi(\boldsymbol{\beta}^*,\boldsymbol{\delta}^*|\boldsymbol{y})=\pi(\boldsymbol{\beta}^*|\boldsymbol{y})\,\pi(\boldsymbol{\delta}^*|\boldsymbol{y},\boldsymbol{\beta}^*),$$

where $\pi(\boldsymbol{\beta}^*|\boldsymbol{y})\approx G^{-1}\sum_{g=1}^{G}\pi\left(\boldsymbol{\beta}^*|\boldsymbol{y},\boldsymbol{z}^{(g)}\right)$, with $\pi\left(\boldsymbol{\beta}^*|\boldsymbol{y},\boldsymbol{z}^{(g)}\right)$ being the full-conditional density for sampling $\boldsymbol{\beta}$ in Step 2 of Algorithm 2 and $\boldsymbol{z}^{(g)}$ are draws from the main MCMC run, so estimation of that ordinate uses random draws that are already available. The ordinate $\pi(\boldsymbol{\delta}^*|\boldsymbol{y},\boldsymbol{\beta}^*)$ is estimated by

$$\pi(\boldsymbol{\delta}^*|\boldsymbol{y},\boldsymbol{\beta}^*)\approx\frac{G^{-1}\sum_{g=1}^{G}\alpha_{MH}\left(\boldsymbol{\delta}^{(g)},\boldsymbol{\delta}^*|\boldsymbol{y},\boldsymbol{\beta}^*\right)q(\boldsymbol{\delta}^*|\boldsymbol{y},\boldsymbol{\beta}^*)}{G^{-1}\sum_{h=1}^{G}\alpha_{MH}\left(\boldsymbol{\delta}^*,\boldsymbol{\delta}^{(h)}|\boldsymbol{y},\boldsymbol{\beta}^*\right)},$$

where the numerator draws $\boldsymbol{\delta}^{(g)}$, $g=1,...,G$, come from $\pi(\boldsymbol{\delta}|\boldsymbol{y},\boldsymbol{\beta}^*)$ and the denominator draws

$\boldsymbol{\delta}^{(h)}$, $h = 1, ..., G$, are draws from $q(\boldsymbol{\delta}|\boldsymbol{y}, \boldsymbol{\beta}^*)$. We note that the latter ordinate is particularly easy to obtain, because $\boldsymbol{\beta}$ is already fixed at $\boldsymbol{\beta}^*$ and $\boldsymbol{z}$ is not involved in the sampling step for $\boldsymbol{\delta}$, so that maximization to determine the moments of $q(\boldsymbol{\delta}|\boldsymbol{y}, \boldsymbol{\beta}^*)$ need only be performed once. Moreover, because the sampling of $\boldsymbol{\delta}$ uses an independence proposal density $q(\boldsymbol{\delta}|\boldsymbol{y}, \boldsymbol{\beta}^*)$ that does not depend on the current value of $\boldsymbol{\delta}$, the draws $\boldsymbol{\delta}^{(h)}$ in the denominator quantity determine the numerator draws $\boldsymbol{\delta}^{(g)}$ once they go through the MH step of the ARMH algorithm and are either accepted or rejected. Therefore, estimation of $\pi(\boldsymbol{\delta}^*|\boldsymbol{y}, \boldsymbol{\beta}^*)$ is done with draws that are obtained concurrently in the same MCMC reduced run.

When univariate models are identified by fixing a second cutpoint but allowing $var(\varepsilon) = \sigma^2$ to be a free parameter to be estimated, the marginal likelihood can be estimated by using the decomposition

$$\pi\left(\boldsymbol{\beta}^*, \sigma^{2*}, \boldsymbol{\delta}^*|\boldsymbol{y}\right) = \pi\left(\boldsymbol{\beta}^*|\boldsymbol{y}\right) \pi\left(\sigma^{2*}|\boldsymbol{y}, \boldsymbol{\beta}^*\right) \pi\left(\boldsymbol{\delta}^*|\boldsymbol{y}, \sigma^{2*}, \boldsymbol{\beta}^*\right),$$

where the first and last ordinates are obtained similarly to the preceding discussion, and the second ordinate $\pi\left(\sigma^{2*}|\boldsymbol{y}, \boldsymbol{\beta}^*\right)$ is estimated as an average $\pi\left(\sigma^{2*}|\boldsymbol{y}, \boldsymbol{\beta}^*\right) \approx G^{-1} \sum_{g=1}^{G} \pi\left(\sigma^{2*}|\boldsymbol{y}, \boldsymbol{\beta}^*, \boldsymbol{z}^{(g)}\right)$ with draws $\boldsymbol{z}^{(g)}$ taken from sampling the distribution $\pi(\boldsymbol{z}, \boldsymbol{\delta}|\boldsymbol{y}, \boldsymbol{\beta}^*)$ in a reduced run given $\boldsymbol{\beta}^*$. For this reason, estimation of the marginal likelihood under the alternative identification restrictions will be more cumbersome except, of course, in cases when there are only $J = 3$ categories so that no cutpoints need to be sampled and this identification scheme may be easy to implement.

Turning attention to multivariate models, we note that estimation of the marginal likelihood follows by straightforward extension of the above methods. Specifically, estimation of the posterior ordinate can be done using the decomposition

$$\pi(\boldsymbol{\beta}^*, \boldsymbol{\Omega}^*, \boldsymbol{\delta}^*|\boldsymbol{y}) = \pi(\boldsymbol{\beta}^*|\boldsymbol{y}) \pi(\boldsymbol{\Omega}^*|\boldsymbol{y}, \boldsymbol{\beta}^*) \pi(\boldsymbol{\delta}^*|\boldsymbol{y}, \boldsymbol{\Omega}^*, \boldsymbol{\beta}^*).$$

Similarly to the previously mentioned cases, estimation of $\pi(\boldsymbol{\beta}^*|\boldsymbol{y})$ is done by averaging the full

conditional density with draws $\left\{\boldsymbol{z}^{(g)}, \boldsymbol{\Omega}^{(g)}\right\} \sim \pi\left(\boldsymbol{z}, \boldsymbol{\Omega}|\boldsymbol{y}\right)$ from the main MCMC run

$$\pi\left(\boldsymbol{\beta}^*|\boldsymbol{y}\right) \approx G^{-1} \sum_{g=1}^{G} \pi\left(\boldsymbol{\beta}^*|\boldsymbol{y}, \boldsymbol{z}^{(g)}, \boldsymbol{\Omega}^{(g)}\right).$$

The next ordinate, $\pi\left(\boldsymbol{\Omega}^*|\boldsymbol{y}, \boldsymbol{\beta}^*\right)$, can be estimated either by

$$\pi\left(\boldsymbol{\Omega}^*|\boldsymbol{y}, \boldsymbol{\beta}^*\right) \approx G^{-1} \sum_{g=1}^{G} \pi\left(\boldsymbol{\Omega}^*|\boldsymbol{y}, \boldsymbol{\beta}^*, \boldsymbol{z}^{(g)}\right)$$

if identification is achieved by fixing two cutpoints for each latent series $\{z_{ik}\}$, $k = 1, ..., q$, so that the full-conditional density of $\boldsymbol{\Omega}$ is inverse Wishart. However, if one instead chooses to pursue estimation under the traditional identification assumption that $\boldsymbol{\Omega}$ is in correlation form with correlations given by $\boldsymbol{\rho}$, then estimation of that ordinate can be done by adapting (13), so that

$$\pi(\boldsymbol{\Omega}^*|\boldsymbol{y}, \boldsymbol{\beta}^*) = \pi(\boldsymbol{\rho}^*|\boldsymbol{y}, \boldsymbol{\beta}^*) = \frac{E_1\left\{\alpha_{MH}\left(\boldsymbol{\rho}, \boldsymbol{\rho}^*|\boldsymbol{y}, \boldsymbol{\beta}^*, \boldsymbol{z}\right) \alpha_{AR}\left(\boldsymbol{\rho}^*|\boldsymbol{y}, \boldsymbol{\rho}, \boldsymbol{z}\right) h\left(\boldsymbol{\rho}^*|\boldsymbol{y}, \boldsymbol{\beta}^*, \boldsymbol{z}\right)\right\}}{E_2\left\{\alpha_{MH}\left(\boldsymbol{\rho}^*, \boldsymbol{\rho}|\boldsymbol{y}, \boldsymbol{\beta}^*, \boldsymbol{z}\right) \alpha_{AR}\left(\boldsymbol{\rho}|\boldsymbol{y}, \boldsymbol{\beta}^*, \boldsymbol{z}\right)\right\}}.$$

The last ordinate $\pi\left(\boldsymbol{\delta}^*|\boldsymbol{y}, \boldsymbol{\Omega}^*, \boldsymbol{\beta}^*\right)$ can be decomposed as $\prod_{k=1}^{q} \pi\left(\boldsymbol{\delta}_k^*|\boldsymbol{y}, \boldsymbol{\Omega}^*, \boldsymbol{\beta}^*, \{\boldsymbol{\delta}_i^* : i < k\}\right)$, where each term is estimated in a reduced run by (12) holding all preceding blocks fixed.

While in the univariate case the quantities in the numerator of (11) are available directly, implementation of these methods in the multivariate case requires the likelihood ordinate $f\left(\boldsymbol{y}|\boldsymbol{\Omega}^*, \boldsymbol{\beta}^*, \boldsymbol{\delta}^*\right)$ in (11), which we obtain by the Geweke, Hajivassiliou, and Keane (GHK) method (Geweke 1991; Börsch-Supan and Hajivassiliou 1993; Keane 1994; Train 2003). In addition, when the normalizing constant of any of the priors is needed for determining the prior ordinate in (11), as may be the case with $\pi\left(\boldsymbol{\rho}\right) \propto N\left(\boldsymbol{\rho}|\boldsymbol{r}_0, \boldsymbol{R}_0\right) 1\left\{\boldsymbol{\rho} \in S\right\}$ when $\boldsymbol{\rho}$ is multivariate, that normalizing constant can be evaluated by simulation, e.g. by drawing $\boldsymbol{\rho} \sim N\left(\boldsymbol{\rho}|\boldsymbol{r}_0, \boldsymbol{R}_0\right)$ and evaluating the frequency with which $\boldsymbol{\rho} \in S$ is satisfied (e.g. Chib and Greenberg 1998).

5 Additional Considerations

5.1 Covariate Effects

In the preceding sections we have presented alternative algorithms for estimating univariate and multivariate ordinal data models. However, interpretation of the resulting parameter estimates is

complicated by the nonlinear and non-monotonic dependence of the response probability on the covariates and the model parameters. To see the possibility for non-monotonicity, consider Figure 1 once again. In that figure, one can see that given the cutpoints, if one increases the mean $\boldsymbol{x}_i'\boldsymbol{\beta}$, the probability of the first category, $\Pr(y_i = 1)$, will decrease and that of the last category, $\Pr(y_i = J)$, will increase. However, those are the only two categories for which the effect of a change in $\boldsymbol{x}_i'\boldsymbol{\beta}$ on the probability of observing that response is monotonic. It is easy to see that for the case given in Figure 1, decreasing $\boldsymbol{x}_i'\boldsymbol{\beta}$ will actually initially increase $\Pr(y_i = 2)$. That probability will reach a maximum when $\boldsymbol{x}_i'\boldsymbol{\beta}$ is decreased to the midpoint between γ_1 and γ_2, while any further decrease in $\boldsymbol{x}_i'\boldsymbol{\beta}$ will cause $\Pr(y_i = 2)$ to fall.

Due to the non-linearity, non-monotonicity, and the alternative identification schemes that can be applied in this setting, we now turn to the question of evaluating the effect of a given covariate x_j on the probability of observing y_i. This is important for understanding the model, for determining the impact of a change in one or more of the covariates, and also for evaluating the plausibility of particular covariate values in setting priors for the model parameters. Because of the nonlinearity of the model, the effect of a change in a covariate depends on all other covariates and model parameters, as well as on the identification restrictions used in estimation. The impact can be quite complex and can be calculated either marginalized over the remaining covariates and the parameters or conditional on some of them, e.g. if we are interested in inference conditional on a particular covariate such as gender, race, or geographical location. Given the specific context, one may consider various scenarios of interest – examples of economic policy interventions may include increasing or decreasing income or taxes by some percentage, requiring an additional year of education, etc. These interventions will affect the probability of response for any one of the ordered categories, but as argued above, that effect can be not only of unknown magnitude, but also of unknown sign, for the intermediate categories.

To illustrate the main ideas, consider the univariate ordinal model

$$z_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i, \quad \text{and} \quad y_i = j \quad \text{if} \quad \gamma_{j-1} < z_i \leq \gamma_j,$$

where we are interested in the effect of a particular x, say x_1, on the $\Pr(y_i = j)$ for some $1 \leq j \leq J$. Splitting $\boldsymbol{x}'_{it}$ and $\boldsymbol{\beta}$ accordingly, we can re-write the above model as

$$z_i = x_{1i}\beta_1 + \boldsymbol{x}'_{2i}\boldsymbol{\beta}_2 + \varepsilon_i.$$

The covariate effect can then be analyzed from a predictive perspective, similarly to thinking about inferences for a hypothetical new individual i. For specificity, suppose that one is interested in the average difference in the implied probabilities between the case when x_{1i} is set to the value $x^{\dagger}_{1i}$ and the case when x_{1i} is set to the value $x^{\ddagger}_{1i}$. Given the values of the other covariates and those of the model parameters $\boldsymbol{\theta}$ (which, in addition to $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, can also include σ^2 depending on the identification restrictions), one can obtain the probabilities $\Pr\left(y_i = j|x^{\dagger}_{1i}, \boldsymbol{x}_{2i}, \boldsymbol{\theta}\right)$ and $\Pr\left(y_i = j|x^{\ddagger}_{1i}, \boldsymbol{x}_{2i}, \boldsymbol{\theta}\right)$, which are available analytically. If one is interested in the distribution of the difference $\left\{\Pr\left(y_i = j|x^{\dagger}_{1i}\right) - \Pr\left(y_i = j|x^{\ddagger}_{1i}\right)\right\}$ marginalized over $\{\boldsymbol{x}_{2i}\}$ and $\boldsymbol{\theta}$ given the data $\boldsymbol{y} = (y_1, ..., y_n)'$, a practical procedure is to marginalize out the covariates using their empirical distribution, while the parameters are integrated out with respect to their posterior distribution. Formally, the goal is to obtain a sample of draws from the distribution

$$\begin{aligned}\left\{\Pr\left(y_i = j|x^{\dagger}_{1i}\right) - \Pr\left(y_i = j|x^{\ddagger}_{1i}\right)\right\} \quad = \quad & \int \left\{\Pr\left(y_i = j|x^{\dagger}_{1i}, \boldsymbol{x}_{2i}, \boldsymbol{\theta}\right) - \Pr\left(y_i = j|x^{\ddagger}_{1i}, \boldsymbol{x}_{2i}, \boldsymbol{\theta}\right)\right\} \\ & \pi(\boldsymbol{x}_{2i})\, \pi(\boldsymbol{\theta}|\boldsymbol{y})\, d\boldsymbol{x}_{2i} d\boldsymbol{\theta}.\end{aligned}$$

A sample from the above predictive distribution can be obtained by the method of composition applied in the following way. Randomly draw an individual and extract the covariate values. Sample a value for $\boldsymbol{\theta}$ from the posterior and evaluate $\left\{\Pr\left(y_i = j|x^{\dagger}_{1i}, \boldsymbol{x}_{2i}, \boldsymbol{\theta}\right) - \Pr\left(y_i = j|x^{\ddagger}_{1i}, \boldsymbol{x}_{2i}, \boldsymbol{\theta}\right)\right\}$. Repeat this for other individuals and other draws from the posterior distribution to obtain draws from the predictive distribution. The mean of that distribution gives the expected difference in the computed pointwise probabilities as $x^{\dagger}_{1i}$ is changed to $x^{\ddagger}_{1i}$, but other quantities such as quantiles and dispersion measures are also easy to obtain given the draws from the predictive distribution.

The above approach can similarly be applied conditionally upon, instead of marginally of, certain variables (such as gender or race) that might determine a particular subsample of interest, in which

case the above procedures are applied only to observations in the subsample. Importantly, the procedures can be applied to multivariate data as well because one only has to consider the marginal distribution for the response of interest. Finally, we note that these techniques can be useful in setting priors on the parameters, where one can calibrate the parameters in the prior distributions by performing the above simulations with draws from the prior (instead of the posterior) in order to see whether specific settings for the hyperparameters produce plausible outcome probabilities and covariate effects.

5.2 Extensions

We now briefly turn attention to some straightforward extensions of the techniques discussed in this paper to other settings. One such extension concerns the analysis of count data because in many cases count data that take a limited number of values can also be viewed and analyzed similarly to ordinal data. For example, in studying a conflict one may be interested in the number of days in a week that are characterized by social unrest or violence, while in transportation economics one may have an interest in the number of vehicles a household owns and uses (Goldberg 1998; West 2002). Fang (2008) uses data from the National Household Transportation Survey to estimate a Bayesian multivariate response system with both discrete (number of vehicles) and censored (miles driven) responses. In labor economics one may be interested in the number of children in a family – e.g. Borg (1989) uses data from the Korean Institute for Family Planning to study the income-fertility relationship using an ordered probit model. In these cases, in order to adapt the ordinal structure to the analysis of count data, there has to be a category that exhausts the set of possible outcomes and is the complement of the categories listed. For this reason, there is usually a response category such as "J or more outcomes" – for example a family can own 1, 2 or "3 or more" cars. When the inclusion of such a remainder category is sensible, ordinal data models can be quite useful in the analysis of count data. This is because they can produce more flexible outcome probabilities over the chosen range of outcomes than those given by

Poisson models since ordinal models are not restricted by the equidispersion property of the Poisson distribution. In multivariate settings, ordinal models can also produce both positive and negative correlation between the responses, whereas the multivariate Poisson model can only accommodate positive correlations; in addition, ordinal models can accommodate both over- and under-dispersion, whereas mixed models such as the Poisson-lognormal model, while capable of capturing negative correlations, can only accommodate overdispersion.

Several other extensions are possible. For instance, the techniques discussed here can be applied in the analysis of longitudinal (panel) data by merging the algorithms provided above with those in Chib and Carlin (1999) for sampling the individual-specific effects. A similar sampler in the context of binary panel outcomes is presented in Chib and Jeliazkov (2006), who consider a semiparametric model with an unknown covariate function. For panel data settings, Algorithm 5 provides an approach for dealing with correlated errors which can also accommodate, in a fairly straightforward fashion, commonly used intertemporal correlation structures, such as exponentially correlated errors, where $\mathbf{\Omega}[t,s] = \exp\{-\alpha|t-s|^r\}$ for scalars α and r (e.g., Diggle and Hutchinson 1989) or various other Toeplitz-type correlation matrices. Yet other extensions can be pursued in the latent variable step; for example, as mentioned earlier, the methods can be adapted to models with mixture of normals or scale mixture of normals link functions (Albert and Chib 1993; Wood and Kohn 1998; Chen and Dey 2000). A non-parametric Bayesian model based on Dirichlet process priors is presented in Kottas, Müller, and Quintana (2005).

6 Applications

Ordinal data outcomes arise frequently in survey data, where respondents may be asked to evaluate a particular issue on an ordinal scale (such as whether they disagree, agree, or strongly agree), as is common in the subjective well-being literature (Duch *et al.* 2000; McBride 2001; Di Tella *et al.* 2003; Luechinger *et al.* 2006). Ordinal outcomes also result when the dependent variable is naturally classified into meaningful ordered categories (e.g. by thresholds in cost or income), such

as working part-time, full-time, or overtime (Kahn and Lang 1996; Olson 1998). Bayesian models involving an endogenous ordinal variable have recently been considered in Li and Tobias (2006) and Herriges *et al.* (2007).

In this section we consider several problems that are of interest in economics and the broader social sciences, and show how the techniques discussed in the preceding sections can be applied in practice. For the univariate case, we consider the widely studied topic of educational attainment using data from the National Longitudinal Study of Youth (NLSY79). Subsequently, we consider two multivariate applications involving survey data on voter opinions and health information. In each application, we estimate the models under the two identification approaches discussed previously – either by fixing two cutpoints or by imposing unit variances – and illustrate and compare the performance of the alternative estimation algorithms by the inefficiency factors for the sampled parameters. The inefficiency factors are calculated as

$$1 + 2\sum_{l=1}^{L} \rho_k(l) \left(\frac{L-l}{L} \right),$$

where $\rho_k(l)$ is the sample autocorrelation for the kth parameter at lag l, and L is chosen at values where the autocorrelations taper off. The inefficiency factors approximate the ratio of the numerical variance of the posterior mean from the MCMC chain relative to that from hypothetical iid draws. The data sets used in our applications can be downloaded from *http://www.econ.uci.edu/~ivan/*.

6.1 Educational Attainment

Educational attainment has been the subject of a large literature that is relevant to researchers and policymakers alike. It is also well suited for empirical study involving ordinal models because the dependent variable is naturally categorized by measurable thresholds of educational attainment such as the completion of high school or college. In our first application the dependent variable, level of education, is measured in four categories: (*i*) less than a high school education, (*ii*) high school degree, (*iii*) some college or associate's degree, and (*iv*) college or graduate degree. In order to facilitate comparability of our results with other research, we estimate a model of educational

attainment using the National Longitudinal Survey of Youth (NLSY79).

Previous research in the area has been abundant. For example, Ermisch and Francesconi (2001) use the British Household Panel Study to estimate the influence of parental education on child educational attainment, while Dearden *et al.* (2002) use the British National Child Development Survey to measure the influence of school quality on educational attainment. Cameron and Heckman (2001) use a dynamic discrete data model and data from the National Longitudinal Survey of Youth to measure the influence of parental income on children's educational attainment. A large literature attempts to estimate the returns to education using a number of methods ranging from explicit proxies for ability, twins studies, fixed effects and arguably exogenous influences on educational attainment. Bayesian studies of the return to schooling include Li and Tobias (2006, 2007), where the ordered probit analysis of the factors that influence educational attainment is embedded into a larger model that studies the subsequent returns to such educational attainment.

A number of studies have analyzed the binary outcome of whether a student graduates from high school (Astone and McLanahan 1991; Haveman, Wolfe and Spaulding 1991; Ribar 1994; Wilson 2001). However, when one is interested in more than just those individuals on the verge of graduating from high school, ordinal data models offer a natural generalization to a choice that is truly among a number of distinct levels of education. As previously noted in Section 5, covariate effects can be rather general in ordinal data models allowing for the theoretically likely possibility that factors such as a student's parental education or income can have varying impacts on educational attainment across categories.

In 1979, the NLSY began annual interviews with over 12000 youths on a battery of demographic questions. Using these data, we estimate the effect of family background, individual, and school variables on educational attainment. The NLSY specifically asks whether the respondent has obtained various education degrees but this information can also be inferred through the years of schooling variable. For this application, we restrict our sample to those cohorts that were ages 14-17 in 1979 for whom a family income variable can be constructed. To create the family income

variable, we average family income over age 16 and 17 when available. The income measure is given in thousands of 1980 dollars. We also restrict our sample by availability of other relevant variables in the data set. Additionally, we exclude disabled individuals and those who report more than 11 years of education at age 15. The resulting sample consists of 3923 individuals. The data set includes variables on an individual's family at the age of 14 including: the highest grade completed by their father and mother, whether the mother worked, family income (stabilized by the square root transformation), and whether the youth lived in an urban area or the South at the age of 14. We also include the individual's gender and race. To control for age cohort affects, we include dummy variables to indicate an individual's age in 1979.

Parameter	Covariate	Mean	SD
β	Intercept	-1.34	0.09
	Family income (sq. rt.)	0.14	0.01
	Mother's education	0.05	0.01
	Father's education	0.07	0.01
	Mother worked	0.03	0.04
	Female	0.16	0.04
	Black	0.15	0.04
	Urban	-0.05	0.04
	South	0.05	0.04
	Age cohort 2	-0.03	0.05
	Age cohort 3	0.00	0.06
	Age cohort 4	0.23	0.06
δ		0.08	0.02
		-0.28	0.03

Table 1: Posterior means and standard deviations for the parameters in the educational attainment application. Identification is achieved through variance restriction and estimation is performed by Algorithm 2 using a sample of 10000 MCMC iterations following a burn-in of 1000 iterations.

The results of our analysis are presented in Table 1. The signs of the coefficients presented in the table are consistent with what is often found in the literature. Parental education as well as income have a positive effect on educational attainment. Labor force participation of the mother has a positive effect on educational attainment. A mother's work force participation could be seen as detrimental due to lack of parental supervision or could be viewed as providing a positive role

model for her children to follow; the sign of the coefficient indicates the latter case is a viable possibility. Conditionally on the remaining covariates, we also see that blacks and individuals from the South have higher educational attainment. To gauge the magnitudes of some of the more interesting covariate effects in this example, one can use the framework presented in Section 5.1. To illustrate these calculations, we computed the effect of an increase in family income of $1000 on educational outcomes. For the overall sample, the effect of such increase in family income is to lower the probability of dropping out of high school by approximately 0.0050, lower the probability of only obtaining a high school degree by 0.0006, but increase the probability of having some college or associate's degree by 0.0020 and increase the probability of getting a college or graduate degree by 0.0036. For the subsample of females, the effects of an income increase on the four outcome probabilities were comparable at approximately -0.0048, -0.0009, 0.0019, and 0.0038, respectively. For the subsample of blacks, the effects of income change were somewhat stronger – in that subsample, an increase of $1000 in family income changed the four educational outcome probabilities by -0.0060, -0.0009, 0.0026, and 0.0043, respectively.

While Table 1 presents results estimated under a variance restriction ($var(\varepsilon) = 1$) by Algorithm 2, we also estimated the model by Algorithm 3 using two types of cutpoint restrictions – $\gamma_1 = 0$ and $\gamma_2 = 1$, and $\gamma_1 = 0$ and $\gamma_3 = 1$, respectively. In the latter two cases $var(\varepsilon) = \sigma^2$ is a free parameter. The point estimates for the parameters from that algorithm, when transformed by $1/\sigma$ produced estimates that were virtually identical to those in Table 1 and were therefore suppressed. However, the inefficiency factors from the three MCMC runs differed, and are presented in Figure 3. The first and second panel in that figure suggest that Algorithm 2 and Algorithm 3 (using transformation (8) due to Chen and Dey (2000)) perform well in this case. A comparison of the second and third panels in Figure 3 shows that identifying the model by fixing the first and last cutpoints is preferable to fixing the first and second cutpoints, indicating that restricting a larger fraction of the latent data $\boldsymbol{z}$ to a fixed range tends to identify the scale better. Regardless of which cutpoints are fixed, however, it is important to keep in mind that Algorithm 3 is a three-block

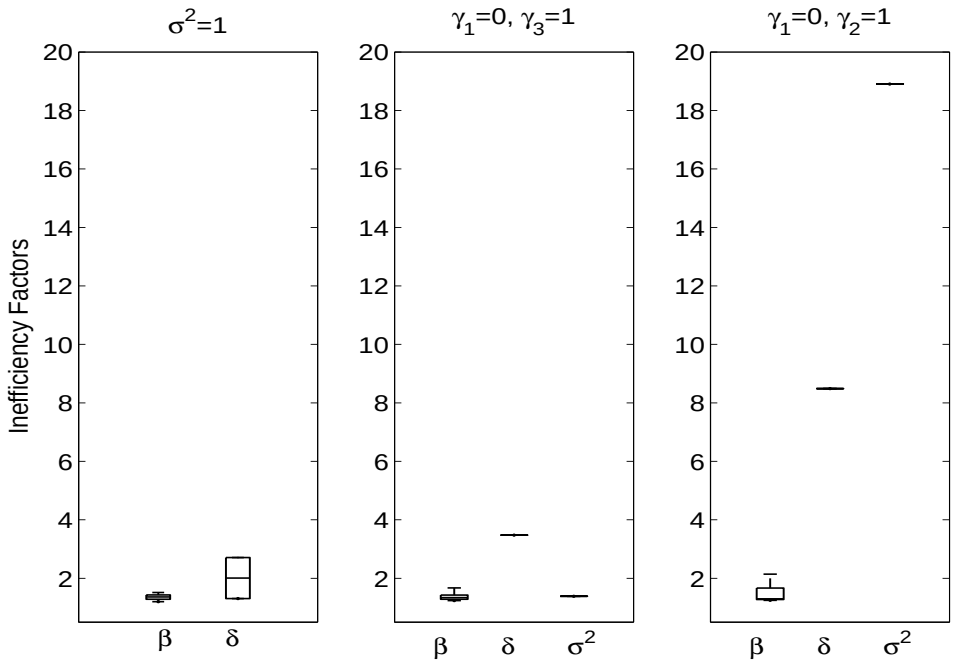

Figure 3: Boxplots of inefficiency factors in the educational attainment application. In the first panel, sampling is implemented under variance restrictions ($\sigma^2 = 1$) and $\boldsymbol{\beta}$ is a 12×1 vector and $\boldsymbol{\delta}$ is 2×1; In the second and third panels, sampling is implemented by Algorithm 3 under cutpoint restrictions, using $(\gamma_1 = 0, \gamma_3 = 1)$ and $(\gamma_1 = 0, \gamma_2 = 1)$, respectively; in those panels, $\boldsymbol{\beta}$ is a 12×1, while δ and σ^2 are scalars.

sampler and requires an additional reduced run when it comes to marginal likelihood estimation relative to Algorithm 2. We note that in all three cases, the MH acceptance rates for $\boldsymbol{\delta}$ were very good and were in the range $(0.90, 0.97)$.

6.2 Voter Opinions

Surrounding election day, political analysts often present descriptive statistics of voter opinions and demographics from exit poll surveys. Such surveys are of particular interest to political economists and politicians since they cast light on voter sentiment and help to predict future election outcomes or guide policy. Recently, an ordered probit model using economic perceptions as the dependent variable was used in Duch, Palmer, and Anderson (2000) to study the degree to which voter perceptions are objective.

When a given survey respondent is asked to comment on multiple issues, his or her responses

will generally be correlated; however, analysts and the media often present the results for various questions independently or as tabulations by demographic groups. In contrast, we estimate the effect of voter demographics on survey responses using a multivariate model that accounts for correlations in a voter's set of responses. We use the National Election Pool General Election Exit Polls, 2006, (Interuniversity Consortium for Political and Social Research, 2007). Our sample consists of 6,050 voters casting a ballot in the 2006 United States general election. Election day and absentee voters were surveyed from October 27, 2006 to November 7, 2006, after leaving polling places or through phone interviews.

The survey attempted to elicit voters opinions on current issues such as how President George W. Bush is performing in office, approval of the war in Iraq, and status of the national economy. Responses to these questions are coded as categorical variables with four ordered categories (higher values imply more favorable opinions). The demographic variables we include are age, sex, race, ethnicity, urban location, region, household income, marital status, and whether children lived in the household. Other explanatory variables include religious affiliation, frequency of religious attendance, and political philosophy. The survey is a stratified random sample that over-samples minorities. The results of fitting this trivariate ordered probit model are presented in Table 2.

The parameter estimates in Table 2 show plausible signs and magnitudes that accord well with intuition. Not surprisingly, relative to moderates, liberals have negative opinions on all three topics, in contrast to conservatives. Respondents in higher income categories, who most likely also have higher education (unavailable in this data set), have a more positive opinion of the national economy, but share a lower opinion of the Iraq war and George W. Bush's performance as President. Those who subjectively consider themselves in a good financial situation have a positive opinion on all three topics. Those in the youngest age category have opinions of smaller magnitude across the board than those in the older age category. Females have a negative opinion of the economy and the Iraq war relative to males, but relatively weak opinion of the President. Coefficients for minorities are negative for all three topics, with larger magnitudes for blacks. Church attenders

		Economy		Iraq War		President	
Parameter	Covariate	Mean	SD	Mean	SD	Mean	SD
β	Intercept	-0.08	0.06	-0.72	0.06	-1.25	0.07
	Liberal	-0.34	0.04	-0.60	0.04	-0.70	0.04
	Conservative	0.49	0.04	0.68	0.04	0.75	0.04
	Income 30k-50k	0.06	0.05	-0.08	0.05	0.00	0.05
	Income 50k-75k	0.08	0.05	-0.13	0.05	-0.04	0.05
	Income 75k up	0.14	0.05	-0.22	0.05	-0.18	0.05
	Financial situation	0.71	0.02	0.48	0.02	0.61	0.02
	Age < 30	0.01	0.05	-0.05	0.05	-0.02	0.05
	Age > 64	0.02	0.04	-0.19	0.04	-0.14	0.04
	Married	0.00	0.03	0.05	0.04	0.05	0.04
	Children	-0.02	0.03	0.04	0.03	0.04	0.03
	Female	-0.21	0.03	-0.09	0.03	0.03	0.03
	Black	-0.44	0.05	-0.53	0.05	-0.67	0.05
	Latino	-0.12	0.06	-0.16	0.06	-0.08	0.06
	Attends church	0.04	0.01	0.10	0.01	0.13	0.01
	Born again	-0.01	0.04	0.18	0.04	0.24	0.04
δ		0.39	0.02	-0.54	0.03	-0.53	0.03
		0.52	0.02	-0.16	0.02	-0.07	0.02

Table 2: Posterior means and standard deviations for the parameters in the voter opinions application. Identification is achieved through variance restrictions and estimates are obtained by Algorithm 5 using a sample of 10000 MCMC iterations after a burn-in of 1000 draws.

(regardless of religion) as well as Born again Christians have a positive view of the Iraq war and President George W. Bush, with Born again Christians having stronger positive opinions.

Because both liberals and conservatives appear strikingly different than moderates (the omitted voter category), we quantify the effect of these two covariates by calculating the implied changes in the response probabilities using the techniques of Section 5.1. To see in practical terms what the effects of the two ideological opposites imply relative to being moderate, we take the subsample of moderates and calculate the response probabilities with and without each dummy. The results are presented in Table 3. We note that a similar exercise can be performed on the subsamples of liberals or conservatives or any other subsample of interest, but one has to be aware that the results need not be identical because the subsamples of respondents will differ in their covariates. Overall, all entries indicate large effects of the liberal and conservative dummies; interstingly, the

	$\Delta \Pr(y_i = 1)$	$\Delta \Pr(y_i = 2)$	$\Delta \Pr(y_i = 3)$	$\Delta \Pr(y_i = 4)$
		Economy		
Moderate → Liberal	0.0734	0.0387	-0.0857	-0.0264
Moderate → Conservative	-0.0699	-0.0963	0.0999	0.0663
		Iraq War		
Moderate → Liberal	0.2148	-0.1364	-0.0756	-0.0028
Moderate → Conservative	-0.2178	0.0450	0.1573	0.0155
		President		
Moderate → Liberal	0.2351	-0.1521	-0.0796	-0.0034
Moderate → Conservative	-0.2323	0.0522	0.1610	0.0191

Table 3: Examples of estimated covariate effects for the parameters of the liberal and conservative dummies in the voter opinion application. The entries indicate the average change in the probability of each outcome for a given covariate change.

magnitudes of these effects are somewhat more balanced for opinions on the national economy than those on the Iraq war and the performance of the President.

Because of the joint modeling and estimation for a voter's set of responses the model takes into account the correlation between voter opinions. The correlation matrix, estimated by Algorithm 5 under unit variance restrictions, is given by

$$\boldsymbol{\Omega} = \begin{pmatrix} 1 & 0.40 & 0.50 \\ 0.40 & 1 & 0.79 \\ 0.50 & 0.79 & 1 \end{pmatrix}.$$

These estimates suggest that the three outcomes in this application are highly positively correlated, which suggests the presence of common unobserved factors influencing all three responses.

In closing, we mention a few additional considerations. As can be expected, the estimates of the parameters from Algorithms 5 and 6 agreed closely after accounting for the different scale identification in the model. However, the inefficiency factors differed very widely as shown in Figure 4, where we see that traditional identification through variance restrictions produces improved mixing of the Markov chain. Moreover, the inefficiency factors presented here are higher than those in the univariate case from the educational attainment application. One reason is that while the vectors of cutpoint differences $\boldsymbol{\delta}_k$ are sampled marginally of the corresponding $\boldsymbol{z}_k$ for $k = 1, \ldots, q$, i.e. $\boldsymbol{\delta}_k \sim \boldsymbol{\delta}_k | \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\rho}, \boldsymbol{z}_{\setminus k}$, the sampling still depends on the latent data $\boldsymbol{z}_{\setminus k}$ for the other responses,

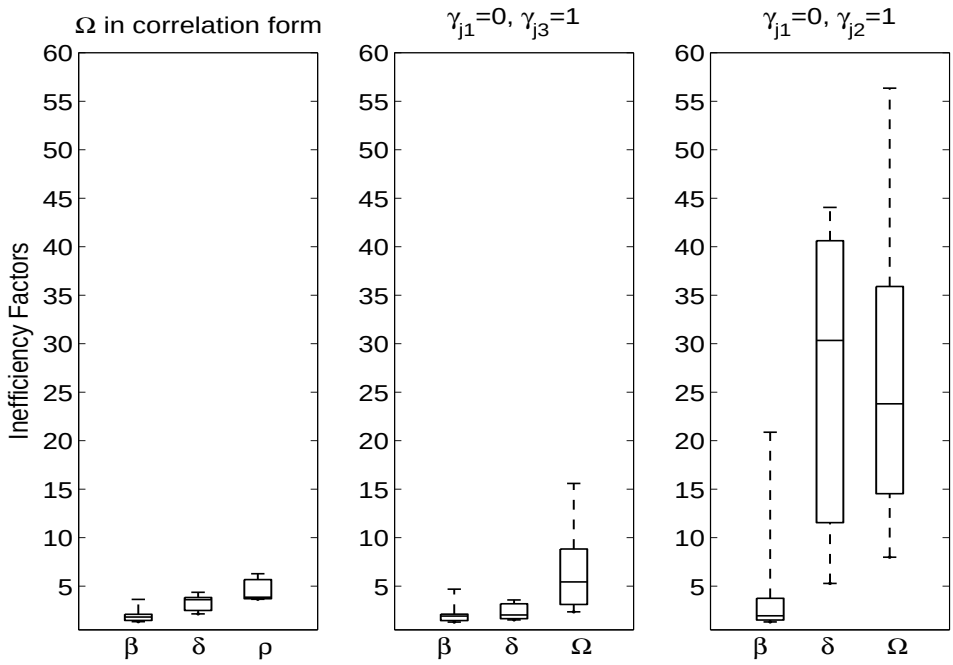

Figure 4: Boxplots of inefficiency factors in the voter opinion application. In the first panel, sampling is implemented under variance restrictions ($\boldsymbol{\Omega}$ is in correlation form with correlations given by $\boldsymbol{\rho}$) by Algorithms 5 and $\boldsymbol{\beta}$ is 48×1, $\boldsymbol{\delta}$ is 6×1, and $\boldsymbol{\rho}$ is 3×1; in the second and third panels, sampling is implemented by Algorithm 6 under cutpoint restrictions, using $(\gamma_{j1} = 0, \gamma_{j3} = 1)$ and $(\gamma_{j1} = 0, \gamma_{j2} = 1)$, respectively; in those panels, $\boldsymbol{\beta}$ is a 48×1, $\boldsymbol{\delta}$ is 3×1, and the unique elements of $\boldsymbol{\Omega}$ form a 6×1 vector.

so that when $\boldsymbol{\Omega}$ exhibits high correlations as is the case in this example, the Markov chain will mix more slowly. This intuition is confirmed in our next example, where the estimated covariance matrix exhibits lower correlations that lead to improved mixing for the parameters. Finally, in both algorithms, the MH steps performed quite well with MH acceptance for $\boldsymbol{\delta}$ in the range $(0.90, 0.95)$ suggesting a close match between the proposal and target densities. Also, by setting the AR constant c in the ARMH algorithm such that $ch\left(\boldsymbol{\rho}|\boldsymbol{z}, \boldsymbol{\beta}\right)/\pi(\boldsymbol{\rho}|\boldsymbol{z}, \boldsymbol{\beta}) = 2$ at the mode of $h\left(\boldsymbol{\theta}\right)$ and setting the ARMH tuning parameter $\tau = 1.5$, in our context we were able to obtain an AR acceptance rate of 0.4 and a corresponding MH acceptance rate of 1 indicating that these settings of the tuning parameters produce an AR sampler for $\boldsymbol{\Omega}$.

6.3 Sources of Medical Information

An essential goal in marketing research is the identification of the source, or sources, of information that consumers use in making decisions to buy. The question is of particular interest when it comes to identifying the sources of medical information, because of the significant investments in drug advertising and the public health consequences of medical choices. Kravitz et. al. (1996) identifies doctors, family and friends, and media advertisements, as the main sources of medical information for most people. Ippolito and Mathios (1990) specifically study the influence of advertisements on health information, while Hellerstein (1998) studies the influence of a physician on patients' use of generic versus brand-name prescription drugs. These methods of obtaining medical information may be correlated. For example, an individual's desire to research a wide array of sources could result in positive correlation between the sources. Alternatively, someone who has a high frequency of obtaining information from a doctor may not rely as often on other sources of information, such as friends and family, or advertising. As a result, it is possible that medical advertising could either substitute for advice from doctors or encourage patients to see a doctor about a concern they have. In our study we use ordered categorical survey responses on the frequency of obtaining information from various sources to address this question.

We use the Public Health Impact of Direct-to-Consumer Advertising of Prescription Drugs, July 2001–January 2002, (ICPSR 2003). The sample consists of entries from 2879 respondents on their self-reported sources of medical information. The sample was created to be nationally representative using random digit dialing, subsequently stratified to over-sample minorities. For our dependent variables we use responses to questions on how often respondents obtain medical information from various sources, namely information obtained from friends and family, through advertisements, and from a doctor. Responses to these questions are coded as categorical variables with four ordered categories indicating frequency. We include additional covariates such as whether the individual has a health condition, medical insurance, or a regular doctor. The demographic variables we include are age, sex, race, ethnicity, education, employment status, student status,

urban location (suburban is the omitted category), region, household income, marital status, and family size.

		Advertisements		Friends & Family		Doctor	
Parameter	Covariate	Mean	SD	Mean	SD	Mean	SD
β	Intercept	0.82	0.13	0.71	0.13	1.44	0.13
	Income 25k to 50k	-0.10	0.05	0.05	0.05	0.06	0.06
	Income 50k to 75k	-0.09	0.07	0.09	0.07	0.07	0.07
	Income 75k up	-0.08	0.07	0.04	0.07	0.11	0.07
	Insurance	0.06	0.07	-0.03	0.07	-0.15	0.07
	Medical Condition	0.02	0.05	-0.03	0.05	-0.37	0.05
	Regular Doctor	0.04	0.06	-0.07	0.06	-0.57	0.06
	Employed	-0.05	0.05	-0.03	0.05	0.15	0.05
	Student	0.07	0.12	-0.13	0.12	-0.38	0.12
	Less than HS	0.22	0.08	0.21	0.08	0.05	0.08
	Some College	0.09	0.06	-0.14	0.06	-0.02	0.06
	College	0.10	0.05	-0.20	0.05	-0.30	0.05
	Age	0.00	0.00	0.01	0.00	0.00	0.00
	Married	-0.04	0.05	-0.08	0.05	-0.14	0.05
	Family size	0.02	0.03	0.01	0.03	0.01	0.03
	Female	-0.15	0.04	-0.18	0.04	-0.18	0.04
	Minority	-0.17	0.05	-0.02	0.05	-0.06	0.05
	Urban	-0.06	0.05	-0.04	0.05	-0.05	0.05
	Rural	-0.06	0.05	0.01	0.05	0.05	0.06
	South	-0.05	0.04	-0.02	0.04	-0.04	0.04
δ		-0.21	0.03	0.04	0.03	0.06	0.03
		-0.40	0.04	-0.28	0.03	-0.16	0.04

Table 4: Posterior means and standard deviations for the parameters in the health information application. Identification is achieved through variance restrictions and estimates are obtained by Algorithm 5 using a sample of 10000 MCMC iterations after a burn-in of 1000 draws.

Parameter estimates obtained by Algorithm 5 are presented in Table 4. The table reveals that relative to those in the lowest income bracket, individuals in higher income brackets are less likely to turn to advertisements as a source of health information, and more likely to turn to friends and family. Higher income brackets are incrementally more likely to obtain health information from a doctor. Surprisingly, those with insurance, a medical condition, or a regular doctor are all less likely to turn to a doctor for medical information. These effects are shown in Table 4, and the covariate effects in Table 5 show the average effect of insurance on the probabilities of the outcome categories

for each source of medical information. One factor that may contribute to some of the surprising

	$\Delta \Pr(y_i = 1)$	$\Delta \Pr(y_i = 2)$	$\Delta \Pr(y_i = 3)$	$\Delta \Pr(y_i = 4)$
		Advertisements		
Insurance = 0 → Insurance = 1	-0.0202	-0.0035	0.0069	0.0168
		Friends & Family		
Insurance = 0 → Insurance = 1	0.0068	0.0031	-0.0014	-0.0085
		Doctor		
Insurance = 0 → Insurance = 1	0.0373	0.0153	-0.0061	-0.0465

Table 5: Estimated covariate effect of insurance. The entries indicate the changes in the probability of each outcome based on the sample of uninsured.

results presented here, is that the variable for medical condition is comprised of a list of serious and well-known medical conditions but does not include minor aches and pains or less serious illnesses. For this reason, many aspects of health which may drive demand for medical information may not be captured in the available data. Additionally, obtaining medical information from a doctor likely requires a higher cost and, unlike with friends and family or advertisements, is more likely to be associated with actually having some physical concern that may not entirely be captured in the documented list of medical conditions. The absence of detailed health information is a limitation of the data.

Because of the joint modeling of the responses, the model accounts for the correlation between health information sources. The correlation matrix, estimated by Algorithm 5 under unit variance restrictions, is given by

$$\mathbf{\Omega} = \begin{pmatrix} 1 & 0.21 & 0.03 \\ 0.21 & 1 & 0.24 \\ 0.03 & 0.24 & 1 \end{pmatrix}.$$

These estimates suggest that certain outcomes in this application are correlated, while others are not. For example, the frequency of using friends and family to obtain medical information is correlated with both information from advertisements and from doctors (0.21 and 0.24, respectively), while the correlation between information from advertisements and doctors is very low (0.03). It may be the case that individuals use friends and family to filter information from the other two sources; this may indicate that information from friends and family serves as a complement to

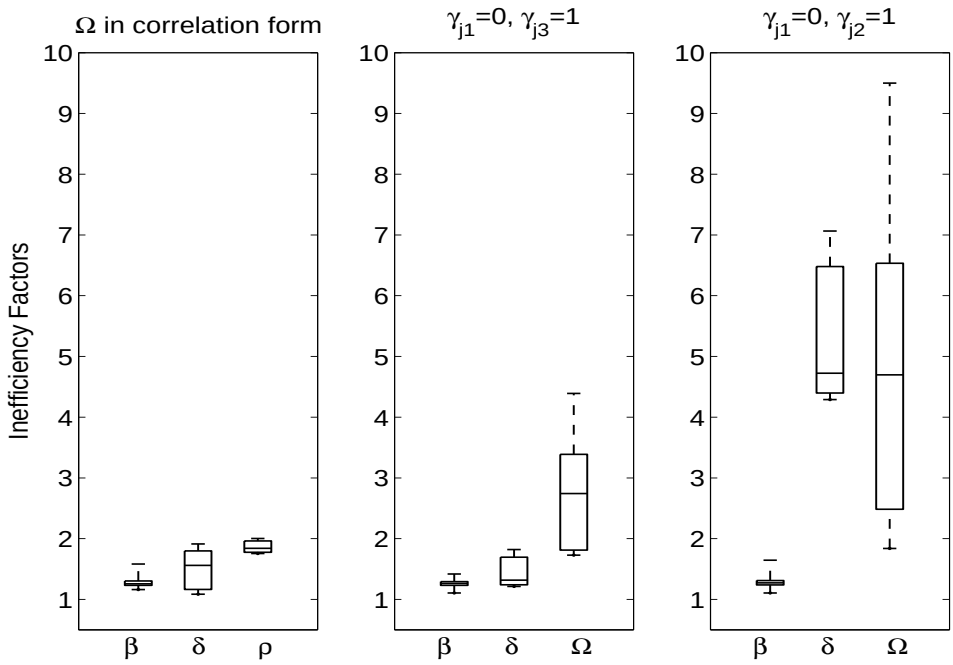

Figure 5: Boxplots of inefficiency factors in the health information application. In the first panel, sampling is implemented under variance restrictions ($\boldsymbol{\Omega}$ is in correlation form with correlations given by $\boldsymbol{\rho}$) by Algorithms 5 and $\boldsymbol{\beta}$ is 60×1, $\boldsymbol{\delta}$ is 6×1, and $\boldsymbol{\rho}$ is 3×1; in the second and third panels, sampling is implemented by Algorithm 6 under cutpoint restrictions, using $(\gamma_{j1} = 0, \gamma_{j3} = 1)$ and $(\gamma_{j1} = 0, \gamma_{j2} = 1)$, respectively; in those panels, $\boldsymbol{\beta}$ is a 60×1, $\boldsymbol{\delta}$ is 3×1, and the unique elements of $\boldsymbol{\Omega}$ form a 6×1 vector.

other sources of medical information. Upon comparing this correlation matrix to the one the voter opinion application, we see that overall the correlations here are much lower.

In Figure 5, we present the inefficiency factors that result from Algorithms 5 and 6 in this application. These overall inefficiency factors appear to be lower when identification is achieved through variance restrictions and estimation is done by Algorithm 5. Again, in both algorithms, the MH steps performed quite well with MH acceptance for $\boldsymbol{\delta}$ again in the range $(0.90, 0.95)$ suggesting a close match between the proposal and the target. As in our voter opinion application, setting the AR constant c in the ARMH algorithm such that $ch\left(\boldsymbol{\rho}|\boldsymbol{z}, \boldsymbol{\beta}\right)/\pi(\boldsymbol{\rho}|\boldsymbol{z}, \boldsymbol{\beta}) = 2$ at the mode of $h\left(\boldsymbol{\theta}\right)$ and the ARMH tuning parameter $\tau = 1.5$, we were able to obtain an AR acceptance rate of 0.41 and a corresponding MH acceptance rate of 1 indicating that the ARMH algorithm essentially involved AR sampling for $\boldsymbol{\Omega}$.

In closing, we return to an interesting difference between the inefficiency factors in this and the previous application. Since sampling of the latent data $\boldsymbol{z}_k$ and cutpoints $\boldsymbol{\gamma}_k$ is conditional on the latent data for the other responses $\boldsymbol{z}_{\backslash k}$ for $k = 1, ..., q$, the magnitude of the correlations in $\boldsymbol{\Omega}$ plays an important role in determining the mixing of the chain. When those correlations are high (as in the voter opinion example), the chain mixes more slowly than when correlations in $\boldsymbol{\Omega}$ are low as in the current example.

7 Conclusion

There are alternative ways in which ordinal models can be identified. In this paper we have discussed some possibilities and shown how they can be implemented in practice using well tailored MH or ARMH algorithms. Our main points can be summarized as follows. First, in the univariate setting, identification through variance restrictions appears to be preferable when the number of categories J is greater than 3. This is because Algorithm 2 allows for more efficient blocking for sampling and marginal likelihood estimation. However, when $J = 3$, a sampler built upon identification by fixing the two cutpoints can be useful as it will not involve any MH steps. Second, in multivariate settings when ordinal models are identified through variance restrictions (leading to a correlation matrix $\boldsymbol{\Omega}$), efficient sampling can be made possible through the ARMH algorithm. In our examples, identification through variance restrictions and estimation through Algorithm 5 was shown to lead to overall improved mixing of the Markov chain relative to identification by cutpoint constraints and fitting by Algorithm 6. Algorithm 5 also allows greater flexibility when $\boldsymbol{\Omega}$ involves restrictions, e.g. when it is structured or involves off-diagonal zeros; however, the easier application of Algorithm 6 can also be appealing when $\boldsymbol{\Omega}$ is not restricted. When identification is achieved by fixing cutpoints, our examples have revealed that fixing the first and last cutpoints (e.g. as in Chen and Dey (2000)) appears to result in lower inefficiency factors than fixing the first and second cutpoints. For these reasons, we recommend Algorithm 5 to more advanced statistical programmers dealing with complex problems, whereas Algorithm 6 can be useful for tackling standard problems

by more customary Markov chains. Third, we have shown that the estimation algorithms discussed here allow for the straightforward calculation of marginal likelihoods and Bayes factors for comparing alternative ordinal models. Finally, the paper has discussed a simulation-based framework for covariate effect evaluation that can be quite useful in eliciting the impact of covariates on the probabilities of ordinal outcomes. The above issues have been illustrated in three important problems in labor economics, political science, and health economics. These studies have demonstrated the applicability and usefulness of the inferential techniques in the context of ordinal data models.

References

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

Albert, J. and Chib, S. (2001). Sequential Ordinal Modeling with Applications to Survival Data. *Biometrics*, 57, 829–836.

Astone, M. and McLanahan, S. S. (1991). Family Structure, Parental Practices and High School Completion. *American Sociological Review*, 56, 309–320.

Borg, M. (1989). The Income-Fertility Relationship: Effect of the Net Price of a Child. *Demography*, 26, 301–310.

Börsch-Supan, A. and Hajivassiliou, V. (1993). Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models. *Journal of Econometrics*, 58, 347–368.

Cameron, S. and Heckman, J. (2001). The Dynamics of Educational Attainment for Black, Hispanic, and White Males. *The Journal of Political Economy*, 109, 455–499.

Chen, M.-H. and Dey, D. K. (2000). Bayesian Analysis for Correlated Ordinal Data Models. In D. Dey, S. Ghosh and B. Mallick (Eds.) *Generalized Linear Models: A Bayesian Perspective* (pp. 133–157). New York: Marcel-Dekker.

Chib, S. (1995). Marginal Likelihood from the Gibbs Output. *Journal of the American Statistical Association*, 90, 1313–1321.

Chib, S. and Carlin, B. (1999). On MCMC Sampling in Hierarchical Longitudinal Models. *Statistics and Computing*, 9, 17–26.

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

Chib, S. and Greenberg, E. (1998). Analysis of Multivariate Probit Models. *Biometrika*, 85, 347–361.

Chib, S. and Jeliazkov, I. (2001). Marginal Likelihood From the Metropolis-Hastings Output. *Journal of the American Statistical Association*, 96, 270–281.

Chib, S. and Jeliazkov, I. (2005). Accept-Reject Metropolis-Hastings Sampling and Marginal Likelihood Estimation. *Statistica Neerlandica*, 59, 30–44.

Chib, S. and Jeliazkov, I. (2006). Inference in Semiparametric Models for Binary Longitudinal Data. *Journal of the American Statistical Association*, 101, 685–700.

Cowles, M. K. (1996). Accelerating Monte Carlo Markov Chain Convergence for Cumulative-link Generalized Linear Models. *Statistics and Computing*, 6, 101–111.

Di Tella, R., MacCulloch, R., and Oswald, A. (2003). The Macroeconomics of Happiness. *Review of Economics and Statistics*, 85, 809–827.

Duch, R., Palmer, H., and Anderson, C. (2000). Heterogeneity in Perceptions of National Economic Conditions. *American Journal of Political Science*, 44, 635–652.

Dearden, L., Ferri, J., and Meghir, C. (2002). The Effect of School Quality on Educational Attainment and Wages. *Review of Economics and Statistics*, 84, 1–20.

Ermisch, J., Francesconi, M. (2001). Family Matters: Impacts of Family Background on Educational Attainments. *Economica*, 68, 137–156.

Fang, H. A. (2008). A Discrete-Continuous Model Of Households' Vehicle Choice and Usage, with an Application to the Effects of Residential Density. *Transportation Research B*, in press.

Gelfand, A. E., and Smith, A. F. M. (1990). Sampling Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, 85, 398–409.

Geweke, J. (1991). Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints. In E. M. Keramidas (Ed.) *Computing Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface* (pp. 571–578). Fairfax VA: Interface Foundation of North America.

Goldberg, P. K. (1998). The Effects of the Corporate Average Fuel Efficiency Standards in the US. *The Journal of Industrial Economics*, 46, 1–33.

Hastings, W. K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika,* 57, 97–109.

Haveman, R., Wolfe, B., and Spaulding, J. (1991). Childhood Events and Circumstances Influencing High School Completion. *Demography*, 28, 133–157.

Hellerstein, J. (1998). The Importance of the Physician in the Generic versus Trade-Name Prescription Decision. *The RAND Journal of Economics*, 29, 108–136.

Herriges, J., Kling, C., Liu, C.-C., and Tobias, J. (2007). What are the Consequences of Consequentiality? Working paper, Department of Economics, Iowa State University.

Interuniversity Consortium for Political and Social Research (2007). National Election Pool General Election Exit Polls, 2006. University of Michigan, Institute for Social Research, study number 3687.

Interuniversity Consortium for Political and Social Research (2007). Public Health Impact of Direct-to-Consumer Advertising of Prescription Drugs, July 2001-January 2002. University of Michigan, Institute for Social Research, study number 4684.

Ippolito, P., and Mathios, A. (1990). Information, Advertising and Health Choices: A Study of the Cereal Market. *The RAND Journal of Economics*, 21, 459–480.

Kahn, S., and Lang, K. (1996). Hours Constraints and the Wage/Hours Locus. *The Canadian Journal of Economics/Revue Canadienne d'Economique*, 29, S71–S75.

Keane, M. (1994). A Computationally Practical Simulation Estimator for Panel Data. *Econometrica*, 62, 95–116.

Kottas, A., Müller, P., and Quintana, F. (2005). Nonparametric Bayesian Modeling for Multivariate Ordinal Data. *Journal of Computational & Graphical Statistics*, 14, 610–625.

Kravitz, R., Callahan, E., Paterniti, D., Antonius, D., Dunham, M., and Lewis, C. (1996). Prevalence and Sources of Patients' Unmet Expectations for Care. *Annals of Internal Medicine*, 125, 730–737.

Li, M., and Tobias, J. (2006). Bayesian Analysis of Structural Effects in an Ordered Equation System. *Studies in Nonlinear Dynamics and Econometrics*.

Li, M., and Tobias, J. (2007). Bayesian Analysis of Treatment Effects in an Ordered Potential Outcomes Model. In D. Millimet, J. Smith and E. Vytlacil (Eds.) *Advances in Econometrics, Volume 21: Estimating and Evaluating Treatment Effects in Econometrics*, in press.

Liu, I., and Agresti, A. (2005). The Analysis of Ordered Categorical Data: An Overview and a Survey of Recent Developments. *TEST*, 14, 1–73.

Luechinger, S., Stutzer, A., and Winkelmann, R. (2006). The Happiness Gains from Sorting and Matching in the Labor Market. *IZA Discussion Papers*, 2019, Institute for the Study of Labor (IZA).

McBride, M. (2001). Relative-Income Effects on Subjective Well-Being in the Cross-Section. *Journal of Economic Behavior and Organization*, 45, 251–278.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equations of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, 21, 1087-1092.

Nandram, B., and Chen, M.-H. (1996). Reparameterizing the Generalized Linear Model to Accelerate Gibbs Sampler Convergence. *Journal of Statistical Computation and Simulation*, 54, 129–144.

O'Brien, S. M., and Dunson, D. B. (2004). Bayesian Multivariate Logistic Regression. *Biometrics*, 60, 739–746.

Olson, C. (1998). A Comparison of the Parametric and Semiparametric Estimates of the Effect of Spousal Health Insurance Coverage on Weekly Hours Worked by Wives. *Journal of Applied Econometrics*, 13, 543–565.

Ribar, D. (1994). Teenage Fertility and High School Completion. *The Review of Economics and Statistics*, 76, 413–424.

Tierney, L. (1994). Markov Chains for Exploring Posterior Distributions. *Annals of Statistics*, 22, 1701–1762.

Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge, UK: Cambridge University Press.

Webb, E. L. and Forster, J. J. (2008). Bayesian Model Determination for Multivariate Ordinal and Binary Data. *Computational Statistics and Data Analysis*, 52, 2632–2649.

West, S. (2002). Distributional Effects of Alternative Vehicle Pollution Control Policies. *Journal of Public Economics*, 88, 735–757.

Wilson, K. (2001). The Determinants of Educational Attainment: Modeling and Estimating the Human Capital Model and Education Production Functions. *Southern Economic Journal*, 67, 518–551.

Wood, S., and Kohn, R. (1998). A Bayesian Approach to Robust Binary Nonparametric Regression. *Journal of the American Statistical Association*, 441, 203–213.

Zellner, A. and Rossi, P. E. (1984). Bayesian Analysis of Dichotomous Quantal Response Models. *Journal of Econometrics*, 25, 365–393.