Estimation and Welfare Analysis in a System of Correlated Count Outcomes[1]

Joseph A. Herriges
Iowa State University

Daniel J. Phaneuf [2]
North Carolina State University

Justin L. Tobias
Iowa State University

January 2008

Abstract

We describe and employ a Bayesian posterior simulator for fitting a high dimensional system of ordinal or count outcome equations. The model is then applied to describe the multiple site recreation demands of individual agents, and we argue that our approach provides advantages relative to existing methods commonly applied in this area. In particular, our model flexibly adjusts to match observed frequencies in trip outcomes, permits a flexible correlation pattern among the sites visited by individuals, and the posterior simulator for fitting this model is relatively easy to implement in practice. We also describe how the posterior simulations produced from the model can be used to conduct a variety of counterfactual experiments, including predicting behavioral changes and describing welfare implications resulting from shifts in exogenous demographic and site characteristics. We illustrate our method using data from the Iowa Lakes Project by modeling the visitation patterns of individuals to a set of twenty-nine large Iowa lakes. Consistent with previous findings in the literature, we see strong evidence that own and cross-price effects on trip demand are negative and positive, respectively, that higher income increases the likelihood of visiting most sites, and that a commonly used indicator of water quality, Secchi transparancy, is positively correlated with the number of trips taken. In addition, the correlation structure among the errors reveals a complex pattern in which unobserved factors affecting trip demand are generally (though not strictly) positively correlated across sites. The flexibility and richness with which we are able to characterize the demand system provides a solid platform for counterfactual analysis, where we find significant behavioral and welfare effects from changes in site availability, water quality, and travel costs.

[1]The authors are listed alphabetically, and not as a reflection of senior authorship. This research was supported in part by the U.S. Environmental Protection Agency and the Iowa Department of Natural Resources. Although the research described in this article has been funded in part by the United States Environmental Protection Agency through R82-5310-010, it has not been subject to the Agency's required peer review policy and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred. All remaining errors are, of course, our own.

[2]Contact author contact information: Box 8109, Department of Agricultural and Resource Economics, North Carolina State University, Raleigh, NC 27695-8109. e-mail: *dan_phaneuf@ncsu.edu*. phone: (919)515-4672.

1 Introduction

Many economic variables take non-negative integer (count) outcomes, and a large literature is devoted to describing econometric models suitable for inference using this type of data. The workhorse models have been based on discrete count distributions such as the Poisson, negative binomial, and their various generalizations that include zero inflation for over dispersion and mixed distributions for unobserved heterogeneity. These approaches have proven effective for modeling univariate outcomes in applications as varied as recreation demand (Englin and Shonkwilder, 1995), cigarette consumption (Mullahy, 1997), and research and development expenditures (Wang et al., 1998). In addition there are several accessible methods for modeling correlated longitudinal (panel) count outcomes; these methods are reviewed by Cameron and Trivedi (1998). Examples of panel count applications are numerous and include airline safety (Dionne et al., 1997), health care utilization (Winkelmann, 2004), and lost workdays (Ruser, 1991). Less well developed, however, are methods suitable for modeling multivariate, correlated count outcomes. Correlated multivariate counts are likely to arise in demand system applications as well as household labor supply models, market entry decisions among oligopolistic firms, and marketing applications. In this paper we employ a modeling approach suitable for this data environment and evaluate its performance with a demand system application examining trips to a set of recreation sites. We also consider how the fitted model can be used for counterfactual analysis of changes in recreation site availability and attributes. Throughout we refer to our model as a Generalized System of Count Outcomes model, or by the convenient acronym GeSCO.

As we discuss below, most existing approaches to modeling correlated multivariate count outcomes start with a basic Poisson or negative binomial distribution for each variable, and then include additional sources of randomness into the specifications for the conditional means. This results in a mixed distribution for each count variable and, if the mixing terms are correlated across the equations, a correlation is induced among the set of count random variables. Since no closed form is available for the marginal (unconditional) probability of observing an outcome, simulation in either its classical form (Hellström, 2006; Egan and Herriges, 2006) or Bayesian form (Chib et al. 1998; Chib and Winkelmann, 2001) is necessary. While approaches exist that do not require simulation (e.g. King, 1989; Winkelmann, 2000) these tend to result in comparatively restrictive characterizations of the correlations between variables.

Though effective in some instances, the reliance on an underlying count distribution such as the Poisson or negative binomial as the host for additional mixing has the potential to be limiting in three important ways. First, these distributions are inherently restrictive in the amount of probability mass that can be accommodated at any one point. This is particularly relevant when the data include many zeros. In applications that do not involve distribution mixing, excess zeros can be accommodated via a variety of zero-inflation approaches. These methods become difficult, however, when they are included in multiple equations requiring simulation for estimation. Second, specification of the conditional mean in standard count distributions is constrained to take the log-linear form. However, in some applications it may be desirable to explore alternative functional forms for the conditional mean, either for consistency with theory in a structural model or to improve model performance. Finally, simulation methods needed for estimation of this class of model are in some instances more challenging than for their non-count counterparts. This is particularly true when the Bayesian paradigm is adopted, in that the conditional posterior distributions of model unknowns often do not take standard forms, thus requiring the use of multiple Metropolis-Hastings steps (and some numerical optimization) to accumulate realizations from the joint posterior.

In contrast to most of the existing literature our approach relies on a strategy in which a multivariate normal distribution maps a set of continuous latent variables into the set of count outcomes that are of interest. In particular, we parameterize the conditional mean of a vector of latent variables that has a normal distribution and general covariance matrix. We then estimate the location of cutoff points in the continuous distribution that map values of the latent variables to values of the discrete outcomes. The use of a normal distribution as our mapping vehicle conveys four advantages. First, we are able to flexibly assign suitable probability mass from the normal distribution to each discrete outcome, thereby avoiding the need for zero inflation or other means of dealing with excess zeros. Second, by allowing a general covariance matrix for the latent variables we induce a quite general pattern of correlation among the count outcomes. Third, we are able to explore a linear functional form for the conditional mean of the latent variable. Finally, if the linear conditional mean is used, estimation from a Bayesian perspective is greatly simplified in that standard forms for most of the conditional distributions needed for posterior simulation are available. Indeed, a Metropolis-Hastings step is only needed in a few instances, and for these, we offer some straight-forward and intuitive suggestions that appear to work well in practice.

We apply our proposed method to a dataset describing visits by Iowa residents to the twenty-nine large lakes in the state. In recreation applications an empirical regularity is that visitation behavior by people is correlated across available sites due to combinations of person-specific avidity and preferences, and the complement/substitute structure of the actual sites. In addition, welfare analysis of changes in site characteristics generally requires use of a demand system approach. Thus recreation applications are an important example of correlated count outcomes. We explain the number of visits people make to each site by including in the latent variable specification the travel cost for accessing the site, prices of other sites in the choice set, personal characteristics of lake visitors, and variables reflecting quality attributes of the lake. We develop a method for predicting behavior under counterfactual scenarios and for predicting the consumer surplus impacts of changes in the availability or quality attributes of the lakes.

Our analysis contributes to several literatures. First, our method provides an additional tool for analyzing an important class of economic variables. Relative to the numerous univariate and panel count applications, examples of correlated multivariate count analysis have been rare and tend to focus on small dimension problems. An important exception in this regard is the recent paper by

Chib et al. (2007), which explicitly considers the model employed here, discusses the merits of alternate restrictions to achieve parameter identification, and introduces and compares a host of alternate approaches for posterior simulation, including a novel use of algorithms that employ an accept-reject Metropolis-Hastings (ARMH) step to facilitate mixing of the posterior simulations. We tailor these methods to issues and problems that arise in recreation demand and similar demand studies and discuss how the simulations produced from such algorithms can be used to conduct welfare and other counterfactual analyses of interest. Within the environmental economics literature, our recreation demand analysis addresses an ongoing issue centered on modeling demand systems that admit interior and corner solutions, accommodate the integer nature of recreation trips, and allow general correlation structures across demand equations. Finally, we contribute to the broader literature on modeling quality-differentiated demands by providing a flexible count-data structure that readily includes non-price determinants of demand and can be applied in situations when multiple choice outcomes are observed.

The remainder of our paper is organized as follows. In the next section we place our GeSCO model in its literature context by reviewing existing approaches to modeling correlated count data, the models for ordinal outcomes upon which our approach is based, and demand system models employing count distributions. In section 3 we present our formal model and discuss how we estimate its unknown parameters from a Bayesian perspective in section 4. Section 5 presents a brief generated data experiment to illustrate the performance of our methods. In section 6 we present methodology regarding counterfactual analysis, and provide an overview of our Iowa lakes application in section 7. Section 8 discusses our actual estimation results and counterfactual simulations and the paper concludes with a summary in section 9.

2 Literature Context

Our modeling approach draws on three areas of literature related to the econometric analysis of discrete outcomes data. We first discuss existing approaches to modeling multivariate, correlated count outcomes and then describe the methods for ordinal data analysis that motives our GeSCO model. We then describe how multivariate count models have been used in demand system analysis when welfare calculation is the objective.

2.1 Correlated Counts

In placing our approach in the context of existing literature on correlated multivariate counts we focus primarily on the methods that are able to provide a general correlation structure across an arbitrary number of count variables; we therefore describe approaches based on mixing distributions. To fix ideas consider the following notation based on Chib and Winkelmann (2001). Let $\boldsymbol{y}_i \equiv$

$(y_{i1}, y_{i2}, \ldots, y_{iJ})'$ denote the collection of J count outcomes for an agent i, where $i = 1, 2, \ldots, I$ is the sample, and for convenience we will refer to the agents as people.[3] Let $\boldsymbol{b}_i \equiv (b_{i1}, b_{i2}, \ldots, b_{iJ})'$ denote a vector of person-specific random terms that are normally distributed such that $\boldsymbol{b}_i \sim N(0, \boldsymbol{D})$ and $\boldsymbol{D}$ is an unrestricted $J \times J$ covariance matrix. Suppose that, conditional on $\boldsymbol{b}_i$ and a vector of unknown parameters $\boldsymbol{\beta}_j$, each element of $\boldsymbol{y}_i$ is distributed independent Poisson:

$$y_{ij}|b_{ij}, \beta_j \overset{iid}{\sim} Poisson(\mu_{ij}), \tag{1}$$

where

$$\mu_{ij} = exp(\boldsymbol{x}'_{ij}\boldsymbol{\beta}_j + b_{ij}), \quad i = 1, 2, \ldots, I; j = 1, 2, \ldots, J, \tag{2}$$

and $\boldsymbol{x}_{ij}$ is a vector of covariates that varies over people, equations, or both. The specification in (1) is the Poisson-lognormal model as described by Aitchison and Ho (1989), and it has two desirable properties. First, the covariance among the elements of $\boldsymbol{b}_i$ induces a general covariance structure between the J count variables. In addition, the potential for over dispersion (the conditional variance greater than the conditional mean) is accommodated so long as the matrix $\boldsymbol{D}$ is positive definite.

The disadvantage of the Poisson-lognormal lies in estimation. Although the distribution of $\boldsymbol{y}_i$ conditional on $\boldsymbol{b}_i$ is a simple product of Poisson probabilities, the unconditional distribution requires integration over the J dimensions of $\boldsymbol{b}_i$:

$$p(\boldsymbol{y}_i|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_J, \boldsymbol{D}) = \int \prod_{j=1}^{J} f_P(y_{ij}|\boldsymbol{\beta}_j, b_{ij})\phi(\boldsymbol{b}_i|\boldsymbol{D})db_{i1}db_{i2}\ldots db_{iJ}, \tag{3}$$

where $f_P(\cdot)$ is the Poisson probability mass function and $\phi(\cdot)$ is the J-dimensional normal probability density function with zero mean and covariance $\boldsymbol{D}$. Thus calculating the probability of observing a person's outcome involves computing a J-dimensional normal integral, and as such simulation will be necessary for problems with more than a few equations.

Given the rapid development of computational methods simulation does not present problems per se; nonetheless there are aspects of this problem that involve particular challenges from both the classical and Bayesian perspectives. Estimation of the unknown parameters $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_J, \boldsymbol{D})$ using classical methods involves simulating (3) for each person in the sample, constructing the simulated sample log-likelihood function, and using a numerical search algorithm to locate the values of the parameters that maximize the simulated log-likelihood. Train (2003) describes contemporary methods for this type of calculation, and also notes problems that can occur in practice. In particular, numerical methods perform poorly when there are flat areas in the likelihood function, and this can easily occur when attempting to estimate a large number of free parameters in $\boldsymbol{D}$. Thus, in practice classical estimation of the Poisson-lognormal model requires ex ante restrictions on $\boldsymbol{D}$ and/or a comparatively small J. While restrictions often make intuitive sense they can be cumbersome to formally test and by their nature reduce the flexibility the model specification is

[3] We follow the convention of using boldface to denote vectors/matrices and reserve capitals strictly for matrices.

designed to exploit. Egan and Herriges (2006) provide an example of a four dimension model describing actual and stated demand for recreation trips, which restricts $\boldsymbol{D}$ to include a single correlation term and four freely estimated variance parameters.

Adopting a Bayesian approach to estimation as in Chib and Winkelmann (2001) addresses many of the difficulties inherent in the classical approach. Maximization of the (simulated) likelihood function is not needed, and the use of prior information (diffuse or otherwise) can facilitate posterior analysis of all parameters in the model, including the unrestricted elements of $\boldsymbol{D}$. As with many complex problems there are therefore advantages to proceeding from a Bayesian perspective. The advantages in this class of problem relative to linear models are somewhat mitigated, however, by the more complex algorithms needed for posterior simulation. As explained by Chib et al. (1998) posterior simulation involves blocking the set of unknown parameters such that values are sequentially sampled from the conditional distributions

$$p(\boldsymbol{b}|\boldsymbol{y}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_J, \boldsymbol{D}); p(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_J, |\boldsymbol{b}, \boldsymbol{D}); p(\boldsymbol{D}|\boldsymbol{b}), \tag{4}$$

where $\boldsymbol{b}$ is a stacked vector holding each $\boldsymbol{b}_i$ for $i = 1, 2, \ldots, I$. Among these conditional distributions $p(\boldsymbol{D}|\boldsymbol{b})$ has a standard form and can be sampled directly, whereas the others require a Metropolis-Hastings step that can be computationally intense under some circumstances. For example, sampling each $p(\boldsymbol{b_i}|\boldsymbol{y}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_J, \boldsymbol{D})$ involves a separate Metropolis-Hastings (M-H) step using a multivariate t distribution for the proposal density, where the mode of the target density is found via an auxiliary numerical optimization step. Likewise sampling from $p(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_J, |\boldsymbol{b}, \boldsymbol{D})$ will in practice involve separate M-H steps for each j, also with an auxiliary numerical step to tune the proposal density. While these posterior sampling algorithms are feasible it is worth investigating the extent to which a similar or greater degree of model flexibly can be maintained while making greater use of conditional distributions that have known forms. Techniques for modeling ordinal data provide a starting point for our alternative approach.

2.2 Ordinal Data

Ordinal data arises in many of the same contexts as count data, and the two data environments share many similarities. Ordinal random variables take discrete values that have meaning in comparison but not in magnitude. A common example is survey questions that ask people to rate the degree to which they agree with a statement; choices often include descriptions such as disagree strongly, disagree, no opinion, agree, and agree strongly. Movement among these outcomes is meaningful, but the common coding device of setting *strongly disagree* = 1, *disagree* = 2,. . . ,*strongly agree* = 5 does not convey quantitative meaning. Formally count outcomes differ in that outcomes do have quantitative meaning. As we argue below, however, this distinction blurs for smaller integer values.

Models for analyzing ordinal data from a Bayesian perspective have been described by Albert and Chib (1993), Chib et al. (2007) and Koop et al. (2007), among others. This class of models involves specification of a latent variable that maps continuous latent values to each ordinal outcome. To illustrate, consider a univariate response variable y_i that can take outcomes indexed by $k = 1, 2, \ldots, K$. Define the latent variable y_i^* such that

$$y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i \quad i = 1, 2, \ldots, I, \tag{5}$$

where $\boldsymbol{x}_i$ is a vector of explanatory variables, $\boldsymbol{\beta}$ is a vector of unknown parameters, and ϵ_i is typically assumed (though not required) to be drawn from a scale-normalized normal distribution. Values of the latent variable are mapped to values for y_i by a threshold crossing rule, given as:

$$y_i = k \quad \text{if} \quad \delta_{k-1} < y_i^* \leq \delta_k. \tag{6}$$

The δ's in the equation above are cutpoints that divide the continuous space over which the latent variable is defined into segments associated with each discrete outcome. Estimating the model involves characterizing the posterior distribution for both $\boldsymbol{\beta}$ and the cutpoints; cutpoints are therefore determined endogenously according to the data and prior information. Thus equations (5) and (6) provide a flexible means of assigning suitable probability mass in the underlying normal distribution to each possible discrete value. Posterior simulation in this model is relatively straightforward in that conventional data augmentation and Gibbs sampling is applied to sequentially sample from the conditional distributions for each y_i^*, β, and the cutpoints. Known distributions are available for the former two pieces, while a relatively straightforward M-H step is typically used for the latter. Multivariate variants of (5) and (6) have recently been described by Chib et al. (2007), and we will ultimately employ similar methods in this paper. Before describing these in detail, however, we turn to describe how the ordinal modeling approach can be adapted to the problem of demand system modeling.

2.3 Demand System Modeling with Count Distributions

Our application considers the demand for visits to a set of recreation sites, and future uses of our proposed econometric model may also involve a demand system setting in which one of the objectives is to measure consumer surplus changes from shifts in price or non-price attributes. To place our welfare calculation approach in context we briefly discuss how count demand systems have been used for welfare analysis in the recreation demand literature.

Suppose interest centers on estimating a system of demand equations for J recreation sites. The typical specification involves setting the conditional mean of a count distribution equal to the expected demand for trips to a recreation site:

$$y_{ij} \overset{iid}{\sim} Poisson(\mu_{ij}), \tag{7}$$

where

$$\mu_{ij} = E(y_{ij}|\boldsymbol{x}_{ij}), \quad i = 1, 2, \ldots, I; j = 1, 2, \ldots, J, \tag{8}$$

y_{ij} is the number of trips taken to site j by person i, $\boldsymbol{x}_{ij}$ is a vector containing prices, income, and other factors thought to influence trip demand, and the conditional mean is interpreted as a demand equation mapping observable characteristics $\boldsymbol{x}_{ij}$ into an expected number of trips. This is the approach taken by Ozuna and Gomez (1994), who estimate a bivariate Poisson model of trips to two recreation sites using a linear specification for expected demand and then compute approximate welfare measures for access to the sites.

More recent applications of count demand system models in recreation analysis (i.e. Englin et al., 1998; von Haefen and Phaneuf, 2003) have interpreted the set of equations for $(\mu_{i1}, \mu_{i2}, \ldots \mu_{iJ})$ as a formal demand system and have imposed restrictions such that the set of expected demands is consistent with a well-defined utility maximization problem (LaFrance, 1992; LaFrance and Hanemann, 1989). These restrictions allow exact, rather than approximate, welfare analysis in the sense that measures of compensating variation rather than consumer surplus are available.[4] A common functional specification is the log-linear ordinary demand system for which

$$E(y_{ij}|p_{ij}, m_i) = exp(\alpha_j + \beta_{jj} p_{ij} + \gamma m_i) \quad i = 1, 2, \ldots, I; j = 1, 2, \ldots, J, \tag{9}$$

where p_{ij} is the travel cost for person i of accessing site j, m_i is income, $(\alpha_j, \beta_{jj}, \gamma)$ are the demand parameters to be estimated, and the intercepts may be functions of site characteristics. In the log-linear specification integrability conditions require all cross price effects are zero, and income effects are constrained to be equal across all the equations. When these restrictions are imposed it is possible to recover the quasi-indirect utility function as

$$V_i = -\frac{1}{\gamma} exp(-\gamma m_i) - \sum_{j=1}^{J} \frac{exp(\alpha_j + \beta_{jj} p_{ij})}{\beta_{jj}} \quad i = 1, 2, \ldots, I, \tag{10}$$

which can be used to compute an exact welfare measure for changes in site access conditions or characteristics. Importantly, equation (10) contains no terms accounting for unobserved heterogeneity and therefore should be interpreted as an expected preference function conditional on observable price and income characteristics. Typically analysts assume that trips follow J independent count distributions with conditional means/expected demands as in (9), and estimate the unknown parameters via maximum likelihood or, in models that include correlation strategies as discussed above, via simulated maximum likelihood. Estimates of the parameters provide a characterization of (10), from which expected welfare measures can be calculated.

Three observations on this strategy for welfare analysis are relevant for our modeling approach. First, the log-linear specification is not ideal for modeling recreation trips in that it includes income effects but not cross price effects, though the latter are much more likely to matter in recreation trip taking behavior. Nonetheless the log-linear specification is favored because it restricts the conditional mean to be strictly positive, as is required for the standard count distributions. Second, welfare measurement is based on expected behavior rather than actual behavior, since preferences

[4] see Hausman, 1981 for historical context.

are based on the conditional means. In this sense the estimation procedure accommodates the integer nature of the data, but post estimation inference is based on a continuous function that does not accommodate the integer nature of trips or reflect unobserved heterogeneity. Finally, efforts to accommodate zero inflation cause further difficulty in welfare measurement, since the extra probability terms require ad hoc decisions on what form the expected demand equations take (see von Haefen and Phaneuf, 2003, for a discussion). These points suggest that exact welfare analysis in systems of count models have historically come at a high cost. We return to this point later when describing our welfare computation algorithms for the GeSCO model in section 6. We now turn to describe the multivariate ordinal outcome model employed in this paper, emphasizing some of its advantages as well as techniques for estimation.

3 The Model

As described in the previous sections of this paper, we seek to employ a model that will remain true to the multivariate, discrete nature of our recreation demand responses and allow for the modeling of a large number of alternative choices. In addition, we want our model to permit a flexible correlation structure among the choices made by individual agents, and at the same time to adapt to the nature of the choice frequencies, such as a large incidence of zero responses. We index agents by $i = 1, 2, \ldots, I$ and sites by $j = 1, 2, \ldots, J$ and presume that the data are balanced - that is, agents completely report trips to all sites under consideration and no data are missing for particular sites.[5]

In addition, for site j we suppose that there are K_j different ordinal responses such that:

$$y_{ij} \in \{0, 1, 2, \ldots, K_j - 1\}.$$

Generally, we can think of y_{ij} as representing a count outcome - the number of trips taken by person i to site j - though this is not strictly necessary. For example, we might lump all trip counts exceeding some threshold into the largest value $K_j - 1$ in the event that, say, a few outlying agents report an unusually large number of trips taken. This interpretation of the y_{ij} values does not affect any estimation steps described in the present section, though it is important to keep in mind in the context of welfare calculations, as will be discussed in section 6. Finally, it is necessary in practice to allow the number of possible ordinal responses to vary across sites such that $K_j \neq K$. Frequently visited sites, for example, will generally require the researcher to employ a larger K_j to adequately model demands for those sites, while the demands for less popular sites may be well-modeled with significantly smaller K_j.

[5] In the context of our application, this is a reasonable assumption as trip data are fully supplied. If this assumption is not appropriate in the context of a different application, an additional step in the sampler can be added to augment the posterior distribution with the missing response data - see Koop, Poirier and Tobias (2007, pp. 246-251) for example.

3.1 Basic Structure

We begin by proposing, like Chib *et al.* (2007), a linear latent variable representation of the process generating trip demand:

$$y_{ij}^* = \alpha_j + \boldsymbol{x}_{ij}\boldsymbol{\beta}_j + \boldsymbol{z}_i\boldsymbol{\gamma}_j + \epsilon_{ij}. \tag{11}$$

In the above, $\boldsymbol{x}_{ij}$ represents a set of characteristics that varies across both individuals and sites. Such variables could include, for example, travel cost to the various locations. In addition, $\boldsymbol{z}_i$ represents characteristics that vary by the individual, but not the site, such as family income, age and other demographic characteristics. The constant, α_j represents a site-specific effect. As discussed later in this section, we will employ a hierarchical prior for the α_j that will enable us to relate these constants to observed characteristics of the site, such as measures of water quality. Finally, the parameters α_j, $\boldsymbol{\beta}_j$, $\boldsymbol{\gamma}_j$ are permitted to vary across sites. This level of generality may not be required and additional restrictions on the parameters can easily be imposed, but we choose to maintain the more general specification in what follows.

Stacking (11) over $j = 1, 2, \ldots, J$ we obtain:

$$\boldsymbol{y}_i^* = \boldsymbol{\alpha} + \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\gamma} + \boldsymbol{\epsilon}_i, \tag{12}$$

where

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_J \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_J \end{bmatrix}, \quad \boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\gamma}_2 \\ \vdots \\ \boldsymbol{\gamma}_J \end{bmatrix},$$

$$\boldsymbol{y}_i^* = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iJ} \end{bmatrix}, \quad \boldsymbol{X}_i = \begin{bmatrix} \boldsymbol{x}_{i1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ 0 & \boldsymbol{x}_{i2} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{x}_{iJ} \end{bmatrix} \quad \text{and} \quad \boldsymbol{Z_i} = \begin{bmatrix} \boldsymbol{z}_i & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ 0 & \boldsymbol{z}_i & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{z}_i \end{bmatrix}.$$

We can then write (12) succinctly as

$$\boldsymbol{y}_i = \boldsymbol{W}_i\boldsymbol{\theta} + \boldsymbol{\epsilon}_i \tag{13}$$

where

$$\boldsymbol{W}_i = [\boldsymbol{I}_J \ \boldsymbol{X}_i \ \boldsymbol{Z}_i], \quad \text{and} \quad \boldsymbol{\theta} = [\boldsymbol{\alpha}' \ \boldsymbol{\beta}' \ \boldsymbol{\gamma}']'.$$

To this point, we have not discussed any distributional assumptions regarding the error vector $\boldsymbol{\epsilon}_i$. Such assumptions, however, are required for a Bayesian analysis, and we assume:

$$\boldsymbol{\epsilon}_i | \boldsymbol{W}_i, \boldsymbol{Z}_i \stackrel{iid}{\sim} \mathcal{N}\left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1J} \\ \rho_{21} & 1 & \cdots & \rho_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{J1} & \rho_{J2} & \cdots & 1 \end{pmatrix} \right] \equiv \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}). \tag{14}$$

In practice, we normalize the conditional variance of each latent variable to unity for identification purposes.[6] In this case, the covariances are interpretable as correlations, and hence the use of ρ_{ij} in the construction of $\mathbf{\Sigma}$ above. Note that the ρ_{ij} are unrestricted in sign - unobservables that make it more likely for an agent to visit a particular site can make it more or less likely that the agent will visit a competing site. Alternate and widely-used models, such as the repeated nested or mixed logit specifications, do not afford this possibility; the structure of the model itself imposes a strong correlation structure and does not exhibit this level of generality (see, for example, Herriges and Phaneuf, 2002). This has significant implications in terms of out-of-sample policy simulations and welfare experiments, as changing the characteristics of one site can have unknown impacts on the trip behavior to other sites.

The model outlined in (11)-(14) is essentially a latent variable SUR model with diagonal restrictions on the covariance matrix. A standard SUR analysis, however, is not appropriate for this and similar applications, as the responses y_{ij} are not continuous. We can, however, think about the y_{ij} as being realizations of a latent linear process, as in (1), where the connection between y_{ij} and y^*_{ij} is given by:

$$y_{ij} = k \quad \text{if} \quad \delta_k^{(j)} < y^*_{ij} \leq \delta_{k+1}^{(j)}, \qquad k = 0, 1, 2, \ldots, K_j - 1, \qquad j = 1, 2, \ldots, J. \tag{15}$$

The $\{\delta_k^{(j)}\}$ are *cutpoints* of the model, and a vector of these are to be estimated for each specific site. Standard identification conditions impose restrictions on some of these cutpoints δ, namely:

$$\delta_0^{(j)} = -\infty, \ \delta_1^{(j)} = 0, \ \delta_{K_j}^{(j)} = \infty \ \forall j. \tag{16}$$

Thus, the vector of unknown cutpoints for alternative j is given as

$$\boldsymbol{\delta}^{(j)} = [\delta_2^{(j)} \ \delta_3^{(j)} \ \cdots \ \delta_{K_j-1}^{(j)}]'. \tag{17}$$

3.2 Hierarchical Priors

To implement a Bayesian analysis, we require prior distributions for the model parameters. We first begin with priors for the site-specific parameters appearing in (11). These are specified as

$$\alpha_j \overset{ind}{\sim} \mathcal{N}(\boldsymbol{q}_j\boldsymbol{\alpha}_0, \sigma^2_\alpha), \quad j = 1, 2, \ldots, J \tag{18}$$

$$\boldsymbol{\beta}_j \overset{iid}{\sim} \mathcal{N}_{k_x}(\boldsymbol{\beta}_0, \mathbf{\Sigma}_{\boldsymbol{\beta}}), \quad j = 1, 2, \ldots, J \tag{19}$$

$$\boldsymbol{\gamma}_j \overset{iid}{\sim} \mathcal{N}_{k_z}(\boldsymbol{\gamma}_0, \mathbf{\Sigma}_{\boldsymbol{\gamma}}), \quad j = 1, 2, \ldots, J, \tag{20}$$

with k_x and k_z denoting the number of columns in X and Z, respectively. In (18) a constant is included in $\boldsymbol{q}_j$ along with other site characteristics, such as indices of water quality or other site

[6] In a thoughtful and detailed paper, Chib et al. (2007) explore alternate algorithms and identification strategies in the context of univariate and multivariate ordinal models. In particular, they develop an accept-reject Metropolis-Hastings (ARMH) algorithm for these types of models that displays some nice mixing properties. We do not employ the ARMH algorithm in the present paper, but instead, make use of a reparameterization as in Nandram and Chen (1996) and Chen and Dey (2000), which is similar to Algorithm 6 of Chib et al. (2007).

amenities that can influence trip demand. In our formulation of the model, these covariates only enter through the distribution of the site-specific constant terms α_j, though, in principle, these could also be added to the distributions of (19) and (20) as well. Note that, marginalized over (18), we obtain a multivariate system where the site characteristics enter directly into the mean function of (11), and a site-specific random effect is also introduced that allows for correlation in outcomes across sites. Finally, the "common means" $\boldsymbol{\alpha}_0$, $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}_0$ are typically of primary interest as a summary of the "overall" effect of the characteristics on aspects of trip demand. Similarly, the variability of these effects across sites, as summarized by σ^2_α, $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$, may also be of interest and policy-relevant.

Priors for the common hyperparameters of (18)-(20) are then added at the terminal stage of the hierarchy:

$$\begin{aligned}
\boldsymbol{\alpha}_0 &\sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \boldsymbol{V}_{\boldsymbol{\alpha}}) && (21)\\
\boldsymbol{\beta}_0 &\sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{V}_{\boldsymbol{\beta}}) && (22)\\
\boldsymbol{\gamma}_0 &\sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\gamma}}, \boldsymbol{V}_{\boldsymbol{\gamma}}) && (23)\\
\sigma^2_\alpha &\sim IG(a_\alpha, b_\alpha) && (24)\\
\boldsymbol{\Sigma}_{\boldsymbol{\beta}}{}^{-1} &\sim W(\rho_\beta \boldsymbol{R}_{\boldsymbol{\beta}}, \rho_\beta) && (25)\\
\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}{}^{-1} &\sim W(\rho_\gamma \boldsymbol{R}_{\boldsymbol{\gamma}}, \rho_\gamma), && (26)
\end{aligned}$$

where all entries on the right-hand side of (21)-(26) are supplied by the researcher. In addition, $IG(\cdot,\cdot)$ represents an *inverse gamma* distribution and $W(\cdot,\cdot)$ denotes a *Wishart* distribution (see, e.g., Koop, Poirier and Tobias (2007, pp. 336, 339, respectively), which are computationally appealing conditionally conjugate prior choices. In practice, values of the terminal parameters above are chosen to be reasonably non-informative so that the data information is dominant. In our empirical work, we set $\boldsymbol{\mu}_{\boldsymbol{\alpha}}$, $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ and $\boldsymbol{\mu}_{\boldsymbol{\gamma}}$ to zero vectors of the appropriate dimensions and $\boldsymbol{V}_{\boldsymbol{\alpha}}$, $\boldsymbol{V}_{\boldsymbol{\beta}}$ and $\boldsymbol{V}_{\boldsymbol{\gamma}}$ to identity matrices of the appropriate dimensions. Similarly, we set $a_\alpha = 3, b_\alpha = 5$ (which sets the prior mean and prior standard deviation of $\sigma^2_\alpha = .1$), $\rho_\beta = k_x + 1, \rho_\gamma = k_z + 1$ and $\boldsymbol{R}_{\boldsymbol{\beta}}$ and $\boldsymbol{R}_{\boldsymbol{\gamma}}$ to identity matrices of the appropriate dimensions. This yields priors that are reasonably non-informative, and whose use seems to perform well in generated data experiments. It is also worth noting that in cases where J is "small," then we have little information directly from the data that can be used to estimate the right-hand side parameters of (18) - (20).[7] In these cases, the priors in (21)-(26) can be quite influential when making posterior statements regarding these common parameters.

Prior distributions must also be added for the covariance matrix $\boldsymbol{\Sigma}$ and cutpoints $\boldsymbol{\delta}$ to complete the model. For the former, the prior for the correlations must be specified jointly to ensure that $\boldsymbol{\Sigma}$ is positive definite. For the latter, a proper prior could be employed that imposes the ordering restriction on the elements of $\boldsymbol{\delta}^{(j)}$, $j = 1, 2, \ldots, J$, but in practice, improper priors are often

[7] That is, one can loosely think of (18)-(20) as regression equations with J observations, though it is important to keep in mind that all parameters will be estimated jointly rather than sequentially.

employed. At this stage we do not explicitly introduce priors for these quantities. Instead, we complete the model by first employing a reparameterization, and then we specify priors over the reparameterized cutpoints and covariance matrix. We now turn to these matters in more detail.

3.3 Reparameterization

We employ the reparameterization strategy of Nandram and Chen (1996), which is further developed by Li and Tobias (2005) for equation systems. This reparameterization is advantageous, as it will simplify aspects of the posterior computations (in some cases significantly so), and has also been shown in related contexts to improve the mixing of the posterior simulations relative to standard Gibbs sampling methods.

To this end, consider taking each y_{ij}^* and multiplying it by the reciprocal of the largest unknown cutpoint. In this regard, let

$$\pi_j = (\delta_{K_j-1}^{(j)})^{-1}, \quad j = 1, 2, \ldots, J,$$

and use the notation $\tilde{\cdot}$ to define the rescaling transformation so that, for example,

$$\tilde{y}_{ij}^* \equiv \pi_j y_{ij}^*, \quad j = 1, 2, \ldots, J, \quad i = 1, 2, \ldots, I. \tag{27}$$

Applying this transformation to (11) gives::

$$\tilde{y}_{ij}^* = \tilde{\alpha}_j + \boldsymbol{x}_{ij}\tilde{\boldsymbol{\beta}}_j + \boldsymbol{z}_i\tilde{\boldsymbol{\gamma}}_j + \tilde{\epsilon}_{ij} \tag{28}$$

or, similar to (12),

$$\begin{aligned}\tilde{\boldsymbol{y}}_i^* &= \tilde{\boldsymbol{\alpha}} + \boldsymbol{X}_i\tilde{\boldsymbol{\beta}} + \boldsymbol{Z}_i\tilde{\boldsymbol{\gamma}} + \tilde{\boldsymbol{\epsilon}}_i \\ &= \boldsymbol{W}_i\tilde{\boldsymbol{\theta}} + \tilde{\boldsymbol{\epsilon}}_i,\end{aligned}$$

with these quantities constructed in the obvious ways, analogous to (11) and (12), using the transformations of (27). Also note that

$$\tilde{\boldsymbol{\epsilon}}_i | \boldsymbol{X}_i, \boldsymbol{Z}_i \stackrel{iid}{\sim} N(0, \tilde{\boldsymbol{\Sigma}}),$$

where

$$\tilde{\boldsymbol{\Sigma}} = \begin{bmatrix} \pi_1^2 & \rho_{12}\pi_1^1\pi_2^1 & \cdots & \rho_{1J}\pi_1^1\pi_J^1 \\ \rho_{21}\pi_1^1\pi_2^1 & \pi_2^2 & \cdots & \rho_{2J}\pi_2^1\pi_J^1 \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{J1}\pi_1^1\pi_J^1 & \rho_{J2}\pi_2^1\pi_J^1 & \cdots & \pi_J^2 \end{bmatrix}. \tag{29}$$

Thus, our transformation has eliminated the diagonal restrictions on $\boldsymbol{\Sigma}$, and replaced these with the square of the reciprocals of the largest free cutpoints. Given this reparameterization, we can now choose to employ a standard, conditionally conjugate prior on $\tilde{\boldsymbol{\Sigma}}$, and to this end, we specify:

$$\tilde{\boldsymbol{\Sigma}} \sim IW\left(\rho_{\Sigma}\boldsymbol{R}_{\boldsymbol{\Sigma}}, \rho_{\Sigma}\right), \tag{30}$$

an inverse Wishart prior with scale matrix $\boldsymbol{R}_{\boldsymbol{\Sigma}}$ and degrees of freedom parameter ρ_{Σ}. In practice, we set $\rho_{\Sigma} = J + 1$ and choose $\boldsymbol{R}_{\boldsymbol{\Sigma}} = \boldsymbol{I}_J$.

The link between the latent data and the observed data is also unchanged by this rescaling transformation. That is,

$$y_{ij} = k \quad \text{if} \quad \tilde{\delta}_k^{(j)} < \tilde{y}_{ij}^* \leq \tilde{\delta}_{k+1}^{(j)}, \quad k = 0, 1, \ldots, K_j - 1, \quad j = 1, 2, \ldots, J.$$

To motivate the value of the reparameterization, consider the potentially realistic case where $K_j = 3 \; \forall j$. In this case, all J cutpoints of the model are "absorbed" in $\tilde{\boldsymbol{\Sigma}}$. That is, when fitting the model, these cutpoints' values can be inferred from the covariance matrix simulations, and no additional steps are required for simulating their values. In other words, in this special case, the posterior simulator is remarkably straight-forward. When $K_j > 3$, however, additional steps will need to be added to sample the remaining cutpoints' values. We turn to handling this more general case in the following section.

3.4 The Augmented Posterior

Let

$$\tilde{\boldsymbol{\Gamma}} = [\tilde{\boldsymbol{\theta}} \;\; \tilde{\boldsymbol{\Sigma}} \;\; \tilde{\boldsymbol{\delta}} \;\; \boldsymbol{\alpha}_0 \;\; \boldsymbol{\beta}_0 \;\; \boldsymbol{\gamma}_0 \;\; \boldsymbol{\Sigma}_{\boldsymbol{\beta}} \;\; \boldsymbol{\Sigma}_{\boldsymbol{\gamma}} \;\; \sigma_{\alpha}^2]$$

denote all parameters of the reparameterized model, noting that

$$\boldsymbol{\pi} = [\pi_1 \;\; \pi_2 \;\; \cdots \;\; \pi_J]$$

is contained in $\tilde{\boldsymbol{\Sigma}}$,

$$\tilde{\boldsymbol{\delta}} \equiv [\tilde{\boldsymbol{\delta}}^{(1)'} \;\; \tilde{\boldsymbol{\delta}}^{(2)'} \;\; \cdots \;\; \tilde{\boldsymbol{\delta}}^{(J)'}]'$$

with

$$\tilde{\boldsymbol{\delta}}^{(j)} = [\tilde{\delta}_2^{(j)} \;\; \tilde{\delta}_3^{(j)} \;\; \cdots \;\; \tilde{\delta}_{K_j-2}^{(j)}].$$

The *augmented* posterior density is the joint posterior distribution of $\tilde{\boldsymbol{\Gamma}}$ and the latent data $\tilde{\boldsymbol{y}}^*$. By Bayes' Theorem, this is obtained as:

$$p(\tilde{\boldsymbol{\Gamma}}, \tilde{\boldsymbol{y}}^* | \mathbf{y}) \propto \left[\prod_{i=1}^{n} \phi(\tilde{\boldsymbol{y}}_i^*; \mathbf{W_i}\tilde{\theta}, \tilde{\boldsymbol{\Sigma}}) \prod_{j=1}^{J} I\left(\tilde{\delta}_{y_{ij}} < \tilde{y}_{ij}^* \leq \tilde{\delta}_{y_{ij}+1}\right) \right] \tag{31}$$

$$\times \left[\prod_{j=1}^{J} p(\tilde{\alpha}_j | \pi_j \boldsymbol{q}_i \boldsymbol{\alpha}_0, \pi_j^2 \sigma_{\alpha}^2) p(\tilde{\boldsymbol{\beta}}_j | \pi_j \boldsymbol{\beta}_0, \pi_j^2 \boldsymbol{\Sigma}_{\boldsymbol{\beta}}) p(\tilde{\boldsymbol{\gamma}}_j | \pi_j \boldsymbol{\gamma}_0, \pi_j^2 \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}) \right] \tag{32}$$

$$\times \; p(\tilde{\boldsymbol{\Sigma}}) p(\boldsymbol{\alpha}_0) p(\boldsymbol{\beta}_0) p(\boldsymbol{\gamma}_0) p(\boldsymbol{\Sigma}_{\boldsymbol{\beta}}) p(\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}) p(\sigma_{\alpha}^2),$$

where a flat prior for the transformed cutpoints $\tilde{\boldsymbol{\delta}}$, $p(\tilde{\boldsymbol{\delta}}) \propto c$ for some constant c, is employed.

The joint posterior for the reparameterized model can be obtained via a change of variables from

$$[\boldsymbol{\Gamma}, \boldsymbol{y}^*], \quad \boldsymbol{\Gamma} \equiv [\boldsymbol{\theta} \;\; \boldsymbol{\Sigma} \;\; \boldsymbol{\delta} \;\; \boldsymbol{\alpha}_0 \;\; \boldsymbol{\beta}_0 \;\; \gamma_0 \;\; \boldsymbol{\Sigma}_{\boldsymbol{\beta}} \;\; \boldsymbol{\Sigma}_{\boldsymbol{\gamma}} \;\; \sigma_\alpha^2]$$

to $[\tilde{\boldsymbol{\Gamma}} \;\; \tilde{\boldsymbol{y}}^*]$. The appearance of the π_j in the right-hand side terms of (32) reflects this change of variable, as the π_j spill over into the mean and variance of the hierarchical prior. A natural question, then, is to ask: since the priors for $\boldsymbol{\Sigma}$ and $\boldsymbol{\delta}$ have not been explicitly specified, what are the implications of the maintained priors for $\tilde{\boldsymbol{\Sigma}}$ and $\tilde{\boldsymbol{\delta}}$ on the priors for the structural quantities?

Some progress in this regard was made by Li and Tobias (2005) in a simplified version of this model. They were able to show that, while the structural priors had rather unusual forms, with suitably chosen hyperparameters they were, in practice, still quite non-informative relative to the data. Essentially, the researcher is faced with a tradeoff that weighs the ease of empirical implementation (i.e., adopting priors in the reparameterized model that combine easily with the augmented likelihood) against coping with rather non-standard implied priors for the structural quantities. In the current paper we regard the former concern as the primary one, and focus on the development and application of a posterior simulator that is reasonably easily implemented, mixes well and fares well in generated data experiments to recover parameters of the data generating process. As shown in the Section 5, and in numerous other experiments that are not reported here, the adoption of these priors seems to perform well in practice, mitigating some of these theoretical concerns.[8]

4 The Posterior Simulator

We fit the model above via a Markov Chain Monte Carlo (MCMC) method that utilizes Gibbs and Metropolis-within Gibbs steps. Many of these are standard, while a few are not. Below, we outline each of the 10 steps that are required to implement the posterior simulator:

Step 1: $\tilde{y}_{if}^*|\cdot, \mathbf{y}$.

From (31), it follows immediately that the posterior conditional for $\tilde{\boldsymbol{y}}_i$ is multivariate truncated normal. As in Geweke (1991), we recognize that the corresponding posterior conditionals for each $\tilde{y}_{ij}$ are *univariate* truncated normal, and exploit this result in order to sample from the desired posterior conditional.

To this end, first let $\tilde{\omega}_{ij}$ denote the (i, j) element of $\tilde{\boldsymbol{\Sigma}}^{-1}$. For brevity in notation, we also define the unconditional mean of $\tilde{y}_{ij}$ as:

$$\tilde{\mu}_{ij} \equiv \tilde{\alpha}_j + \boldsymbol{x}_{ij}\tilde{\boldsymbol{\beta}}_j + \boldsymbol{z}_i\tilde{\boldsymbol{\gamma}}_j \tag{33}$$

[8]Another concern regarding this prior is that it does not impose any ordering or range restrictions on the elements of $\tilde{\boldsymbol{\delta}}$. In our view, this is an insignificant concern, as these properties will be enforced through our choice of proposal density for sampling $\tilde{\boldsymbol{\delta}}$, as described in Step 4 below.

and the conditional mean of $\tilde{y}_{ij}$ (given the other elements of $\tilde{\boldsymbol{y}}_i$) as:

$$\tilde{\mu}_{i-j} \equiv \tilde{\mu}_{ij} - \tilde{\omega}_{jj}^{-1} \sum_{k \neq j} \tilde{\omega}_{ik} (\tilde{y}^*_{ik} - \tilde{\mu}_{ik}). \tag{34}$$

For a given individual i, we can then independently sample, for $j = 1, 2, \ldots, J$:

$$\tilde{y}^*_{ij} | \cdot, \mathbf{y} \sim TN_{(\tilde{\delta}_{y_{ij}}, \tilde{\delta}_{y_{ij}+1}]}(\tilde{\mu}_{i-j}, \tilde{\omega}_{jj}^{-1}), \tag{35}$$

where $x \sim TN_{[a,b]}(\mu, \sigma^2)$ denotes that x has a normal distribution with mean μ and variance σ^2 which is truncated to the interval $[a, b]$. In this case, and in all those that follow, the "$\cdot$" in the conditioning implies that we condition on all parameters other than the parameter being sampled. This process is then repeated for $i = 1, 2, \ldots, I$.

Step 2: $\tilde{\boldsymbol{\theta}} | \cdot, \mathbf{y}$.

Using the result of Lindley and Smith (1972), the posterior conditional for the vector of parameters $\tilde{\boldsymbol{\theta}}$ is:

$$\tilde{\boldsymbol{\theta}} | \cdot, \mathbf{y} \sim \mathcal{N}(D_{\tilde{\boldsymbol{\theta}}} d_{\tilde{\boldsymbol{\theta}}}, D_{\tilde{\boldsymbol{\theta}}}), \tag{36}$$

where

$$D_{\tilde{\boldsymbol{\theta}}} \equiv \left[\sum_{i=1}^{n} \boldsymbol{W}_i' \tilde{\boldsymbol{\Sigma}}^{-1} \boldsymbol{W}_i + \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\theta}}}^{-1} \right]^{-1} \qquad d_{\tilde{\boldsymbol{\theta}}} \equiv \sum_{i=1}^{n} \boldsymbol{W}_i' \tilde{\boldsymbol{\Sigma}}^{-1} \tilde{\boldsymbol{y}}^*_i + \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\theta}}}^{-1} \boldsymbol{\mu}_{\tilde{\boldsymbol{\theta}}}.$$

The terms $\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\theta}}}$ and $\boldsymbol{\mu}_{\tilde{\boldsymbol{\theta}}}$ denote the implied prior mean and covariance matrix for $\tilde{\boldsymbol{\theta}}$ from (21)-(23). To characterize these precisely, let us first introduce some additional notation. Let

$$\mathbf{H} \equiv diag(\{\pi_j\}) \quad \text{and define } \boldsymbol{H}_{j,k} \equiv \pi_j \boldsymbol{I}_k,$$

where $\pi_j = [\delta^{(j)}_{K_j - 1}]^{-1}$ has been previously defined. With this notation in hand, it follows that:

$$\boldsymbol{\mu}_{\tilde{\boldsymbol{\theta}}} \equiv \begin{bmatrix} \boldsymbol{H}\boldsymbol{Q}\boldsymbol{\alpha}_0 \\ \boldsymbol{H}_{1,k_x}\boldsymbol{\beta}_0 \\ \vdots \\ \boldsymbol{H}_{J,k_x}\boldsymbol{\beta}_0 \\ \boldsymbol{H}_{1,k_z}\boldsymbol{\gamma}_0 \\ \vdots \\ \boldsymbol{H}_{J,k_z}\boldsymbol{\gamma}_0 \end{bmatrix}$$

and

$$\boldsymbol{\Sigma}_{\tilde{\boldsymbol{\theta}}} \equiv \begin{bmatrix} \sigma_\alpha^2 \boldsymbol{H}\boldsymbol{H}' & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \boldsymbol{H}_{1,k_x}\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\boldsymbol{H}_{1,k_x} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \boldsymbol{H}_{J,k_x}\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\boldsymbol{H}_{J,k_x} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \boldsymbol{H}_{1,k_z}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\boldsymbol{H}_{1,k_z} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \boldsymbol{H}_{J,k_z}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\boldsymbol{H}_{J,k_z} \end{bmatrix}.$$

Step 3: $\tilde{\boldsymbol{\Sigma}}|\cdot, \mathbf{y}$.

From (31), it follows that the posterior conditional for $\tilde{\boldsymbol{\Sigma}}$ is

$$p(\tilde{\boldsymbol{\Sigma}}|\cdot, \mathbf{y}) \propto p(\tilde{\boldsymbol{\Sigma}}) \left[\prod_{i=1}^{n} \phi(\tilde{\boldsymbol{y}}_i^*; \boldsymbol{W}_i\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\Sigma}})\right] \left[\prod_{j=1}^{J} p(\tilde{\alpha}_j; \pi_j \boldsymbol{\alpha}_0, \pi_j^2 \sigma_\alpha^2) p(\tilde{\boldsymbol{\beta}}_j | \pi_j \boldsymbol{\beta}_0, \pi_j^2 \boldsymbol{\Sigma}_{\boldsymbol{\beta}}) p(\tilde{\boldsymbol{\gamma}}_j | \pi_j \boldsymbol{\gamma}_0, \pi_j^2 \boldsymbol{\Sigma}_{\boldsymbol{\gamma}})\right]. \tag{37}$$

The second group of terms in (37) emerge from the change of variables from $\boldsymbol{\theta}$ to $\tilde{\boldsymbol{\theta}}$, and these terms clearly involve the diagonal elements of $\tilde{\boldsymbol{\Sigma}}$.

To discuss how to simulate draws from (37), first note that the prior $p(\tilde{\boldsymbol{\Sigma}})$ and augmented likelihood combine naturally to yield an

$$IW\left(\left[(\rho_\Sigma \boldsymbol{R}_{\boldsymbol{\Sigma}})^{-1} + \sum_{i=1}^{I} (\tilde{\boldsymbol{y}}_i^* - \boldsymbol{W}_i\tilde{\boldsymbol{\theta}})(\tilde{\boldsymbol{y}}_i^* - \boldsymbol{W}_i\tilde{\boldsymbol{\theta}})'\right)^{-1}, \rho_\Sigma + J\right) \tag{38}$$

density.

We choose (38) as a proposal density and implement a Metropolis-Hastings step to sample $\tilde{\boldsymbol{\Sigma}}$. Given this choice, we first sample a $\tilde{\boldsymbol{\Sigma}}^*$ from (38) and then accept it with probability:

$$\min\left\{1, \frac{\prod_{j=1}^{J} p(\tilde{\alpha}_j; \pi_j^* \boldsymbol{q}_j \boldsymbol{\alpha}_0, [\pi_j^*]^2 \sigma_\alpha^2) p(\tilde{\boldsymbol{\beta}}_j | \pi_j^* \boldsymbol{\beta}_0, [\pi_j^*]^2 \boldsymbol{\Sigma}_{\boldsymbol{\beta}}) p(\tilde{\boldsymbol{\gamma}}_j | \pi_j^* \boldsymbol{\gamma}_0, [\pi_j^*]^2 \boldsymbol{\Sigma}_{\boldsymbol{\gamma}})}{\prod_{j=1}^{J} p(\tilde{\alpha}_j; \pi_j \boldsymbol{q}_j \boldsymbol{\alpha}_0, \pi_j^2 \sigma_\alpha^2) p(\tilde{\boldsymbol{\beta}}_j | \pi_j \boldsymbol{\beta}_0, \pi_j^2 \boldsymbol{\Sigma}_{\boldsymbol{\beta}}) p(\tilde{\boldsymbol{\gamma}}_j | \pi_j \boldsymbol{\gamma}_0, \pi_j^2 \boldsymbol{\Sigma}_{\boldsymbol{\gamma}})}\right\}.$$

In practice, virtually all the candidates are accepted, as expected, since the proposal density is a close match to the target, and differs from the target only through contributions of the hierarchical prior to the conditional posterior distribution.

Step 4: $\tilde{\boldsymbol{\delta}}|\cdot, \mathbf{y}$.

We sample the cutpoints within our model separately by site, and employ a random-walk type Metropolis-Hastings step to accomplish this. When $K_j > 3$, the unknown (and rescaled) cutpoint vector is:[9]

$$\tilde{\boldsymbol{\delta}}_j = [\tilde{\delta}_2^{(j)} \quad \tilde{\delta}_3^{(j)} \quad \cdots \quad \tilde{\delta}_{K_j-2}^{(j)}].$$

[9]Note that this step is not needed when $K_j \leq 3$ since, in these cases, there are either no unknown cutpoints, or their values are embedded in $\tilde{\boldsymbol{\Sigma}}$ upon reparameterization.

Marginalized over the latent data for site j we obtain the target posterior conditional distribution:[10]

$$p(\tilde{\boldsymbol{\delta}}_j|\cdot,\mathbf{y}) \propto \prod_{i=1}^{n} \Phi\left((\tilde{\delta}_{y_{ij}+1} - \tilde{\mu}_{i-j})\tilde{\omega}_{jj}^{1/2}\right) - \Phi\left((\tilde{\delta}_{y_{ij}} - \tilde{\mu}_{i-j})\tilde{\omega}_{jj}^{1/2}\right) \tag{39}$$

where the relevant terms in (39) are defined in Step (1).

Since this distribution does not take a standard form, we implement a Metropolis-Hastings step, choosing a proposal density similar to the recommendation of Cowles (1996). To describe this choice of proposal density, let $\tilde{\boldsymbol{\delta}}_*^{(j)}$ denote a candidate value sampled from this density, which can potentially depend on the current value in the chain, which is denoted as $\tilde{\boldsymbol{\delta}}^{(j)}$. We choose to employ a proposal density of the form

$$p(\tilde{\boldsymbol{\delta}}_*^{(j)}|\tilde{\boldsymbol{\delta}}^{(j)}) = p(\tilde{\delta}_{2,*}^{(j)}|\tilde{\boldsymbol{\delta}}^{(j)})p(\tilde{\delta}_{3,*}^{(j)}|\tilde{\delta}_{2,*}^{(j)},\tilde{\boldsymbol{\delta}}^{(j)})\cdots p(\tilde{\delta}_{K_j-2,*}^{(j)}|\tilde{\delta}_{K_j-3,*}^{(j)},\tilde{\boldsymbol{\delta}}^{(j)}). \tag{40}$$

That is, we have broken the joint proposal density into a "marginal" for the smallest cutpoint value, $\tilde{\delta}_{2,*}^{(j)}$ times the corresponding sequence of conditional distributions. In these conditional distributions, we only require dependence on the last value drawn from the proposal density and make this explicit in the notation above. Each of these pieces in (40), of course, can depend on the current value of the chain, $\tilde{\boldsymbol{\delta}}^{(j)}$, which is also made explicit in the notation. We choose

$$\tilde{\delta}_{l,*}^{(j)}|\tilde{\boldsymbol{\delta}}^{(j)} \sim TN_{(\tilde{\delta}_{l-1,*}^{(j)},\tilde{\delta}_{l+1}^{(j)}]}(\tilde{\delta}_l^{(j)},d^2), \quad l = 2,3,\ldots,K_j-2. \tag{41}$$

To explain the intuition behind this proposal, we begin by sampling $\tilde{\delta}_{2,*}^{(j)}$ from a normal distribution which is centered at the current value of the chain $\tilde{\delta}_l^{(j)}$ with variance d^2 and is truncated to the interval $(0,\tilde{\delta}_3^{(j)})$. The variance parameter d^2 can be tuned to yield reasonable acceptance rates - values that are too large will tend to produce few accepted candidates, while values that are too small will yield very small movements from consecutive iterations.

Once $\tilde{\delta}_{2,*}^{(j)}$ is sampled, we then draw sample $\tilde{\delta}_{3,*}^{(j)}$ from a normal distribution centered at the current value of the chain with variance d^2 which is truncated to the interval $(\tilde{\delta}_{2,*}^{(j)},\tilde{\delta}_4^{(j)})$. Thus, our proposal density clearly enforces the ordering restriction on the cutpoints values, as each consecutive element of $\tilde{\boldsymbol{\delta}}_*^{(j)}$ must exceed the previous value. This process is repeated until the last cutpoint, $\tilde{\delta}_{K_j-2,*}^{(j)}$ is sampled. We then accept the vector

$$\tilde{\boldsymbol{\delta}}_*^{(j)} = [\tilde{\delta}_{2,*}^{(j)} \quad \tilde{\delta}_{3,*}^{(j)} \quad \cdots \quad \tilde{\delta}_{K_j-2,*}^{(j)}]$$

with probability

$$p = \min\left\{1, \frac{p(\tilde{\boldsymbol{\delta}}_*^{(j)}|\cdot,\mathbf{y})}{p(\tilde{\boldsymbol{\delta}}_*^{(j)}|\tilde{\boldsymbol{\delta}}^{(j)})}\frac{p(\tilde{\boldsymbol{\delta}}^{(j)}|\tilde{\boldsymbol{\delta}}_*^{(j)})}{p(\tilde{\boldsymbol{\delta}}^{(j)}|\cdot,\mathbf{y})}\right\}.$$

[10]In standard ordered probit analyses, it is well-known that the standard Gibbs sampler, which samples from the latent data given the cutpoints and then the cutpoints given the latent data, mixes very poorly. To mitigate this poor mixing, blocking steps are often employed where the cutpoints and latent variables are sampled together. We follow the spirit of this approach in integrating out the latent data from the j^{th} equation when sampling $\tilde{\boldsymbol{\delta}}_j$.

For the ratio of proposal ordinates in the expression for p above, note that the normal kernels in the truncated normal densities will cancel in the ratio, though it is necessary to calculate the normalizing constant of each term in (40) at both the current and candidate values. When there is substantial variability in K_j across sites, a more refined algorithm can be employed which tailors the choice of tuning parameter d^2 for each j, though we have found that setting $d^2 = .001$ to work reasonably well in generated data experiments. Finally, note that $y_{ij} = 0$ or $y_{ij} = K_j - 1$ outcomes can be ignored when evaluating (39), as the likelihood contributions associated with these choices do not involve any unknown cutpoints after reparameterization. This process is repeated for $j = 1, 2, \ldots, J$, and in this way, a vector of cutpoints is simulated for each site at each iteration in the sampler.

Step 5: $\boldsymbol{\alpha}_0|\cdot, \mathbf{y}$.

The remaining steps of our posterior simulator are rather standard. Below, since we condition on $\boldsymbol{\pi}$ (since it is known given $\tilde{\boldsymbol{\Sigma}}$), we can first "untransform" the parameter vectors and work directly with the "structural" parameters as given in (18) - (26). Thus, in the posterior conditionals below, the $\tilde{}$ notation is dropped; the transformation is primarily as a convenient device for sampling $\tilde{\boldsymbol{\Sigma}}$, and once that is done, the sampling of the other parameters can proceed in a largely familiar way. For the common mean of the site-specific intercept parameters, for example,

$$\boldsymbol{\alpha}_0|\cdot, \mathbf{y} \stackrel{ind}{\sim} \mathcal{N}(D_{\boldsymbol{\alpha}_0} d_{\boldsymbol{\alpha}_0}, D_{\boldsymbol{\alpha}_0}),$$

where

$$D_{\boldsymbol{\alpha}_0} \equiv (\boldsymbol{Q'Q}/\boldsymbol{\sigma}^2_{\boldsymbol{\alpha}} + \boldsymbol{V}^{-\mathbf{1}}_{\boldsymbol{\alpha}})^{-1}, \quad \text{and} \quad d_{\boldsymbol{\alpha}_0} \equiv \boldsymbol{Q'\alpha}/\sigma^2_\alpha + \boldsymbol{V}^{-\mathbf{1}}_{\boldsymbol{\alpha}} \boldsymbol{\mu}_{\boldsymbol{\alpha}}.$$

Step 6: $\sigma^2_\alpha|\cdot, \mathbf{y}$.

$$\sigma^2_\alpha|\cdot, \mathbf{y} \sim IG\left[\frac{n + a_\alpha}{2}, \left(b_\alpha^{-1} + [1/2]\sum_{j=1}^{J}(\alpha_j - \boldsymbol{q}_j\boldsymbol{\alpha}_0)^2\right)^{-1}\right].$$

Step 7: $\boldsymbol{\beta}_0|\cdot, \mathbf{y}$.

$$\boldsymbol{\beta}_0|\cdot, \mathbf{y} \stackrel{ind}{\sim} \mathcal{N}(D_{\boldsymbol{\beta}_0} d_{\boldsymbol{\beta}_0}, D_{\boldsymbol{\beta}_0}),$$

where

$$D_{\boldsymbol{\beta}_0} \equiv (J\boldsymbol{\Sigma}^{-\mathbf{1}}_{\boldsymbol{\beta}} + \boldsymbol{V}^{-\mathbf{1}}_{\boldsymbol{\beta}})^{-1}, \quad \text{and} \quad d_{\boldsymbol{\beta}_0} \equiv \boldsymbol{\Sigma}^{-\mathbf{1}}_{\boldsymbol{\beta}} \sum_{j=1}^{J} \boldsymbol{\beta}_j + \boldsymbol{V}^{-\mathbf{1}}_{\boldsymbol{\beta}} \boldsymbol{\mu}_{\boldsymbol{\beta}}.$$

Step 8: $\boldsymbol{\Sigma}^{-\mathbf{1}}_{\boldsymbol{\beta}}|\cdot, \mathbf{y}$.

$$\boldsymbol{\Sigma}^{-\mathbf{1}}_{\boldsymbol{\beta}}|\cdot, \mathbf{y} \sim W\left(\left[(\rho_\beta \boldsymbol{R}_{\boldsymbol{\beta}})^{-1} + \sum_{j=1}^{J}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_0)(\boldsymbol{\beta}_j - \boldsymbol{\beta}_0)'\right]^{-1}, J + \rho_\beta\right).$$

Step 9: $\boldsymbol{\gamma_0}|\cdot, \mathbf{y}$.

$$\gamma_0|\cdot, \mathbf{y} \overset{ind}{\sim} \mathcal{N}(D_{\gamma_0} d_{\gamma_0}, D_{\gamma_0}),$$

where

$$D_{\gamma_0} \equiv (J\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} + \boldsymbol{V}_{\boldsymbol{\gamma}}^{-1})^{-1}, \quad \text{and} \quad d_{\gamma_0} \equiv \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} \sum_{j=1}^{J} \boldsymbol{\gamma}_j + \boldsymbol{V}_{\boldsymbol{\gamma}}^{-1} \boldsymbol{\mu}_{\boldsymbol{\gamma}}.$$

Step 10: $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1}|\cdot, \mathbf{y}$.

$$\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1}|\cdot, \mathbf{y} \sim W\left(\left[(\rho_\gamma \boldsymbol{R}_{\boldsymbol{\gamma}})^{-1} + \sum_{j=1}^{J} (\boldsymbol{\gamma}_j - \gamma_0)(\boldsymbol{\gamma}_j - \gamma_0)' \right]^{-1}, J + \rho_\gamma \right).$$

A posterior simulator proceeds by successively drawing from the densities outlined in Steps 1 - 10.

5 A generated data experiment

In this section, we introduce a generated data experiment. This experiment is conducted to illustrate the ability of the algorithm to uncover parameters of the data generating process and also to illustrate the mixing properties of the posterior simulations. To fix ideas, we implement this method in a small setting where $J = 5$ and and K_j varies across sites. Specifically, we set $K_j = 3, 5, 5, 4, 6$ for $j = 1, 2, 3, 4, 5$, respectively. Thus, after reparameterization, the number of free cutpoints ranges from 0 (when $K_j = 3$) to 3 (when $K_j = 6$). However, even in the $K_j = 3$ case, it is important to note that one cutpoint is still recovered from the model; its value is simply obtained through the posterior simulations of the transformed covariance matrix and no additional sampling steps are required until $j > 3$. In these cases where such sampling is required, we employ the method outlined in step 4 above to draw the remaining cutpoints' values.

In addition to the above, we set $I = 2,000$. The variables in $\boldsymbol{Q}$ consist of an intercept and a randomly generated standard normal random variable, the variables in $\boldsymbol{X}$ consist of a standard normal and a uniform random variable, while the variables in $\boldsymbol{Z}$ consist of two independent standard normal random variables. To generate the data, we begin with the terminal stage of the hierarchy and sample α_j, $\boldsymbol{\beta}_j$ and $\boldsymbol{\gamma}_j$ from (18)-(20), respectively for $j = 1, 2, \ldots, 5$. In this process, we specify values for $\boldsymbol{\alpha}_0$, $\boldsymbol{\beta}_0$ γ_0 (which are provided in Table 1 below) and set $\sigma_\alpha^2 = .01$, $\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = .01 I_{k_x}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} = .01 I k_z$. These site-specific intercepts and slopes are then employed in (11) to generate the latent y_{ij}^*. In these calculations, we specify

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & .2 & .2 & 0 & 0 \\ .2 & 1 & -.3 & 0 & 0 \\ .2 & -.3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{42}$$

Finally, the latent vector of values is then mapped into a observed ordinal response vector via (15), where the cutpoint values are specified in Table 1 below. To fix ideas on the performance of our method, the cutpoints and other parameters of our data generating process were chosen to yield a reasonably well-balanced distribution of responses for each possible site. In the site with the largest number of ordinal outcomes, for example (i.e., $j = 5$ where $K_5 = 6$), the least-represented category had approximately 9 percent of the outcomes (181/2000), while the most represented category held approximately 35 percent of the outcomes (697/2000). In instances where the data are not as well-balanced across ordinal values, we would expect significantly less posterior precision and perhaps slower mixing regarding some parameters of interest, particularly the cutpoints which serve to define the probabilities associated with the infrequent responses.

We run the Gibbs sampler for 1,000 iterations and discard the first 250 of these as the burn-in period. The sampler is run in the transformed parameter space, as outlined in section 4, and for each post-convergence simulation, we then transform back to obtain the "structural" parameters of the model in (11)-(15). A choice of 1,000 simulations is conservative to be sure, but it is our intent to illustrate here that the simulator converges very quickly to the parameters of the data generating process and performs well even with this small number of simulations when the responses are reasonably balanced. Finally, we start the chain at values that are clearly away from the true parameters of the data generating process; all of the regression parameters in (11) are set to 0, all covariance matrices are set equal to identity matrices of the appropriate dimensions, $\sigma_\alpha^2 = .5$, and the re-scaled cutpoints are set to values between 0 and 1 that do not align with their true values.

In Table 1 we present coefficient posterior means and their true values used in the data generation process. We find that the posterior means are quite close to the actual values used to generate the data. In cases where there is some divergence between the mean and actual values, the difference between these values typically lies within a posterior standard deviation of the parameter. The posterior standard deviations, however, are not reported in the table for the sake of brevity. It is worth noting nonetheless that there is a large amount of posterior uncertainty surrounding the common means given in the final block of the table. In this illustrative run, we have $J = 5$ so that there is little information coming from the data about the values of this "population" distribution. In larger applications with more sites, we can obtain more precise estimates of these common parameters. It is reassuring, however, that even with $J = 5$, the posterior means of these parameters are quite close to the generated data values.

The posterior mean of the covariance matrix $\boldsymbol{\Sigma}$ in this application is given as follows:

$$E(\boldsymbol{\Sigma}|\boldsymbol{y}) = \begin{bmatrix} 1 & .23 & .16 & -.04 & -.00 \\ .23 & 1 & -.34 & -.00 & -.00 \\ .16 & -.34 & 1 & -.01 & -.02 \\ -.04 & -.00 & -.01 & 1 & -.04 \\ -.00 & -.00 & -.02 & -.04 & 1 \end{bmatrix},$$

which, again, is quite close to the values used in (42).

Finally, we provide in Figure 1 some graphical evidence regarding the mixing of our posterior simulations. Specifically, we select two representative cutpoints, a particular α_j, elements of β_0 and γ_j and the hierarchical variance parameter σ^2_α, and examine the lagged autocorrelations among these simulations. If the lagged autocorrelations are quite high, then the sampler does not move much from iteration to iteration, resulting in estimates with high numerical standard deviations, and potentially misleading inference if an adequate number of simulations are not taken. As shown in Figure 1, however, the mixing of the posterior simulations is remarkably good. The "worst" cases occur with the mixing of the cutpoints. The lag-1 autocorrelation among these values is around .7, but these decay very quickly. In fact, the lag-10 autocorrelations are virtually zero. For the other parameters, these virtually resemble iid sampling, as the autocorrelations decay to zero within a few lag orders.

Another way to look at the mixing of the posterior simulations is to compare how many draws are needed to achieve the same level of numerical precision that would be obtained under *iid* sampling. These so-called *inefficiency factors* can be quite useful, and can be calculated by noting that the numerical standard error of a Monte Carlo estimate with correlated draws can be obtained as:

$$NSE(\overline{\eta}) = \sqrt{\frac{\sigma^2}{m}\left[1 + 2\sum_{j=1}^{m-1}\left(1 - \frac{j}{m}\right)\rho_j\right]},$$

where ρ_j is a measure of the correlation between draws j iterations apart, $\sigma^2 = \text{Var}(\eta)$, and m represents the number of post-convergence simulations. Since $\rho_j > 0$ generally, the numerical standard error of the Gibbs estimate exceeds $\sigma/\sqrt{m}$, the numerical standard error obtained under iid sampling.

When calculating these inefficiency factors (defined as the ratio of NSE above over $\sigma/\sqrt{m},$) for a selection of parameters, we found that none of these exceeded 2, and most were within the range of 1.4-1.7. Thus, in order to achieve the same level of accuracy as one would get with m iid draws, we need to generate no more than $1.7m$ posterior simulations. This, again, suggests that our algorithm mixes quite well in this experiment.

6 Counterfactual Analysis

The algorithm described in Section 4 provides a characterization of the joint posterior distribution for all the unknowns in the model, including parameters of the latent demand equations, the correlation structure among the equations, and the set of cutpoints for each demand equation. In this section we discuss how the estimated model can be used for counterfactual analysis, such as predicting behavioral changes and consumer surplus measures associated with exogenous shifts in characteristics of the modeled goods. In the particular case of recreation demand we are often interested in three measures: predicted changes in total visits to sites, changes in the probability of participation (i.e. switching between visitor and non-visitor status), and the welfare implications associated with changes in site attributes or other exogenous conditions. We define each of these measures in turn and outline a general algorithm for their computation.

In some cases, quantities of interest center on the number of trips made to a *single* site. We begin by describing procedures for carrying out these types of calculations, though we will later generalize this to the case of multiple sites. Regardless of scope, the Bayesian considers all of these efforts as problems of *posterior prediction*. In other words, we seek to use the given data and what we have learned from that data about parameters of the model to say something about hypothetical trip patterns under policy scenarios that have not yet been observed. To this end, we introduce the notation T^s_{fj} to denote the number of trips taken by some representative future agent f to site j under scenario s.[11] By "scenario," what we have in mind is to change the covariates' values in some particular way and to follow how this change does (or does not) impact recreation demand. The characteristics that can be manipulated in our model consist of $\boldsymbol{q}^s_j, \boldsymbol{x}^s_{fj}$ and $\boldsymbol{z}^s_f$. It might be of interest, for example, to start with a baseline $\boldsymbol{q}^0_j$ and consider improving water quality to a level of $\boldsymbol{q}^1_j$ and seeing how this change impacts visitation patterns to a particular site. Alternatively, we can consider the case of an increase in travel cost to a given site and thus define an analogous change in the characteristics $\boldsymbol{x}^s_{fj}$. Regardless of how these changes are defined, any such experiment implies a resulting change in trip demand, as represented by the difference $T^1_{fj} - T^0_{fj}$.

Let $\boldsymbol{\Gamma}$ denote all parameters of the model, and for brevity in notation, let us simply use $\boldsymbol{W}_f$ in the conditioning notation to denote the covariates' values given under scenarios $s = 0$ and $s = 1$. We then seek to obtain the posterior predictive distribution

$$p(T^1_{fj} - T^0_{fj} | \boldsymbol{W}_f, \mathbf{y}).$$

Note that this distribution remains a function of the covariates' values assigned in the counterfactual experiment. This dependence, however, is not particularly limiting. In practice, we will simulate from the posterior predictive above for each agent in the sample, and then use the empirical distribution of the sample characteristics to average the individual-level gains or losses into a final summary impact.[12]

To describe how this is done we consider, without loss of generality, the posterior predictive distri-

[11]We distinguish at this point between the indicator variable, y_{ij}, that identifies which of K_j categories (or bins) that an individual belongs to and the number of trips actual taken by that individual. This distinction is important if one or more of the categories corresponds to a range of trips possibly taken by individuals in that category. In our application below, this is the case for the upper-most category, which aggregates all individuals taking more than $K_j - 2$ trips (i.e.,$y_{ij} = K_j - 1 \Rightarrow T_{ij} \geq K_j - 1$)

[12]That is, the baseline characteristics (denoted with the 0 superscript) represent the observed characteristics for the individual. The manipulated characteristics (with a 1 superscript) will then replace $\boldsymbol{q}^0_j$ with a new value, or change travel cost in $\boldsymbol{x}^0_{fj}$ by a given percentage for each agent. Posterior predictive means are then calculated for each agent, and finally averaged across agents to produce an overall impact measure.

bution for a particular agent. This is defined as follows:[13]

$$
\begin{aligned}
p(T^1_{fj} - T^0_{fj} | \boldsymbol{W}_f, \mathbf{y}) &= \int\int\int p(T^1_{fj} - T^0_{fj}, y^1_{fj}, y^0_{fj}, \boldsymbol{\Gamma} | \boldsymbol{W}_f, \mathbf{y}) \; d\boldsymbol{\Gamma} \; dy^1_{fj} \; dy^0_{yf} \\
&= \int\int\int p(T^1_{fj} - T^0_{fj} | y^1_{fj}, y^0_{fj}, \boldsymbol{W}_f, \boldsymbol{\Gamma}, \mathbf{y}) p(y^1_{fj}, y^0_{fj} | \boldsymbol{\Gamma}, \boldsymbol{W}_f, \mathbf{y}) p(\boldsymbol{\Gamma} | \mathbf{y}) \; d\boldsymbol{\Gamma} \; dy^1_{fj} \; dy^0_{yf}.
\end{aligned}
$$

The method of decomposition suggests that if we can draw from $p(\boldsymbol{\Gamma}|\mathbf{y})$ (which is the output from the posterior simulator), $p(y^1_{fj}, y^0_{fj}|\boldsymbol{\Gamma}, \boldsymbol{W_f}, \mathbf{y})$, and then $p(T^1_{fj} - T^0_{fj}|y^1_{fj}, y^0_{fj}, \boldsymbol{W_f}, \boldsymbol{\Gamma}, \mathbf{y})$,we can draw directly from the desired posterior predictive.

For the last of these densities (i.e., the first density of the triple integral above), if the ordinal values created in the modeling process align perfectly with the number of trips taken (which would be the case if the data are well-balanced), then drawing from this distribution is trivial. That is, if $y^s_{fj} = k$ then $T^s_{fj} = k,\ s = 0, 1$. Thus, the distribution is degenerate given y^s_{fj} and the difference is easily calculated as $T^1_{fs} - T^0_{fs} = y^1_{fs} - y^0_{fs}$. In the context of our application, however, when $y^s_{fj} = K_j - 1$, this only indicates that $T^s_{fj} \geq K_j - 1$. That is, the largest category accounts for all of trip outcomes equal to or exceeding $K_j - 1$. The model's aggregations of individuals in this last category will in general preclude calculating trip behavior of individuals within this group.

In practice, when $y^s_{fj} = K_j - 1$ we simply set $T^s_{fj} = K_j - 1$ as a bound. When our experiment considers an enhancement in site conditions, this will generally lead to an understatement of the increase in the number of trips taken, whereas an experiment involving a degradation of characteristics will generally lead to an understatement (in absolute terms) of the decrease in the number of trips taken. Thus, our posterior results under this assumption can be interpreted as a bounded estimate of the distributional impact.[14] The impact of this assumption on posterior predictive simulations depends, of course, on the distribution of households amongst and within the individual bins. For many of the alternatives in our application, the probabilities associated with the largest bins are small, so that simulated $y^s_{fj} = K_j - 1$ outcomes are quite unlikely under realistic experiments and the contribution of these terms will play a small role in the corresponding posterior predictive simulations. Though not pursued here, one could also ameliorate this issue by considering a finer ordinal outcome vector that more completely categorizes the discrete response trip data by creating further subdivisions of the largest bin. In such a setting, however, the prior will play a more influential role in the posterior calculations.

To complete the composition step and provide a way to draw from the desired posterior predictive, we still require a method of drawing from $p(y^1_{fj}, y^0_{fj}|\boldsymbol{\Gamma}, \boldsymbol{W}_f, \mathbf{y})$. Below, we describe three steps for obtaining a draw from this distribution, adding a "(r)" subscript on the components of $\boldsymbol{\Gamma}$ to clearly denote how this process changes with each post-convergence simulation.

[13] Note that the posterior $p(\boldsymbol{\Gamma}|\mathbf{y})$ is independent of the future charactersitics $\boldsymbol{W}_f$.

[14] Alternatively, the assumption should not be considered restrictive if one believes that the most zealous users of the sites are not affected by changes in the site characteristics, in which case the difference in trips taken can be set at zero for all those in the highest bin.

Step 1: Draw $\boldsymbol{\alpha}^s_{(r)}$ from (18). That is, for $j = 1, 2, \ldots, J$, set

$$\alpha^s_{j,(r)} = \boldsymbol{q}^s_j \boldsymbol{\alpha}_{0(r)} + \sigma_{\alpha(r)} \eta_{j(r)}, \quad s = 0, 1$$

where $\eta_{j(r)} \sim N(0, 1)$. Importantly, note that (r) indexes *posterior* simulations of the common mean $\boldsymbol{\alpha}_0$ and hierarchical variance σ^2_α.

Step 2:

Draws of the latent demand vector $\boldsymbol{y}^{*s}_{f(r)} | \boldsymbol{\Gamma}_{(r)}, \boldsymbol{W}_f$ are obtained using

$$\boldsymbol{y}^{*s}_{f(r)} = \boldsymbol{\alpha}^s_{(r)} + \boldsymbol{X}^s_f \boldsymbol{\beta}_{(r)} + \boldsymbol{Z}^s_f \boldsymbol{\gamma}_{(r)} + \boldsymbol{\epsilon}_{f(r)} \quad s = 0, 1, \tag{43}$$

where

$$\boldsymbol{\epsilon}_{f(r)} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{(r)}). \tag{44}$$

The same error draw $\epsilon_{f(r)}$ is used to calculate both $\boldsymbol{y}^{*1}_{f(r)}$ and $\boldsymbol{y}^{*0}_{f(r)}$, embodying our implicit assumption that these errors capture unobservable attributes associated with the individual that are not impacted by the changes under the counterfactual scenario.

Step 3:

The latent demand vector from step 1 can then be mapped into the corresponding ordinal outcome vector $\boldsymbol{y}^s_{f(r)}$ using the threshold crossing rule:

$$y^s_{fj(r)} = k \quad \text{if} \quad \delta^{(j)}_{k(r)} < y^{*s}_{fj(r)} \le \delta^{(j)}_{k+1(r)}, \qquad k = 0, 1, 2, \ldots, K_j - 1;\ j = 1, 2, \ldots, J, \quad s = 0, 1. \tag{45}$$

To summarize, draws from the posterior predictive $p(T^1_{fj} - T^0_{fj} | \boldsymbol{W}_f, \mathbf{y})$ are obtained as follows. First, a value of $\boldsymbol{\Gamma}$ is taken from our posterior simulator. Second, this value is used to generate a pair of values y^1_{fj} and y^0_{fj} for a particular site j following Steps 1-3 above. Finally, these ordinal values are then mapped into observed trip counts, noting that this mapping is somewhat arbitrary for the case of the largest bins in our data. This generates a draw from the desired posterior predictive distribution for a given individual under the policy scenario. This process can then be repeated for all agents in the sample to get an overall, (e.g., mean) treatment impact.

The above posterior predictive is rather narrow in the sense that it ignores the linkages among the sites. In a multiple site model it is, perhaps, more interesting to simultaneously consider the change in visits to all sites, allowing for substitution and complementary effects to be taken into account. The steps given above readily accommodate this more general prediction. The simultaneous approach is particularly relevant if one is interested in the total trips taken by households or the propensity of households to opt out of recreational activity altogether. In these instances, simulation based methods provide the simplest means of computing the impact of changing conditions. For example, one area of interest centers on how recreation visitors change their participation status in response to exogenous attribute shifts. This is an extensive margin summary that gauges the probability that a person will exit or enter the market under changed circumstances conditional on having participated or not participated under baseline conditions. In particular, we may be interested in

$$\Xi(\boldsymbol{q}_j^0, \boldsymbol{q}_j^1, \boldsymbol{X}_f^0, \boldsymbol{X}_f^1, \boldsymbol{Z}_f^0, \boldsymbol{Z}_f^1) \equiv Pr\left(\boldsymbol{y}_{fj}^1 = 0\ \forall j, \sum_{j=1}^{J} \boldsymbol{y}_{fj}^0 \neq 0 \big| \mathbf{y}\right). \tag{46}$$

A simulation-based estimate of Ξ follows naturally as:

$$\hat{\Xi}(\boldsymbol{q}_j^0, \boldsymbol{q}_j^1, \boldsymbol{X}_f^0, \boldsymbol{X}_f^1, \boldsymbol{Z}_f^0, \boldsymbol{Z}_f^1) = \frac{1}{R}\sum_{r=1}^{R} I\left(\sum_{j=1}^{J} \boldsymbol{y}_{fj(r)}^1 = 0\right) I\left(\sum_{j=1}^{J} \boldsymbol{y}_{fj(r)}^0 \neq 0\right), \tag{47}$$

where $I(\cdot)$ denotes the indicator function, and the values $y_{fj(r)}^s$ are obtained as in Steps 1-3 above. Different choices of the covariates values define different experiments, as is clearly represented by the dependence of Ξ on $\boldsymbol{q}_j^s, \boldsymbol{X}_f^s$ and $\boldsymbol{Z}_f^s$ in (46).

The final area of counterfactual interest lies in welfare measurement. Unlike the count system models described in the previous section we do not attempt exact welfare measurement since we do not at this stage impose the theoretical demand system restrictions. Thus our approach to welfare measurement relies on consumer surplus approximations. Our approach does, however, explicitly account for unobserved heterogeneity in a manner that is not available in the typical count demand system welfare measurement. In addition our specification allows for a rich accommodation of both substitution and income effects and thus will reflect more behavioral responses than has typically been done in count system welfare analysis.

To describe our welfare measurement algorithm, consider Figure 2, which depicts a step-wise demand curve of some representative future agent f for site j. The terms along the vertical axis, $c_{fj}^{k,s}$, are essentially reservation prices that (holding all else equal) define the boundaries between consuming k and *k-1* units of good j; i.e., $P_{fj}^s \geq c_{fj}^{k,s} \Rightarrow y_{fj}^s < k$. The individual reservation prices are readily defined using the latent demand equations and the cutpoints as:

$$c_{fj}^{k,s}(\boldsymbol{\Gamma}, \boldsymbol{W}_f) = \frac{\delta_{k+1}^{(j)} - [\alpha_j^s + \boldsymbol{x}_{-p,fj}^s \boldsymbol{\beta}_{-p,j} + \boldsymbol{z}_f^s \boldsymbol{\gamma}_j + \varepsilon_{fj}]}{\beta_{pj}}, \tag{48}$$

where $\boldsymbol{x}_{-p,fj}^s$ denotes the vector of covariates for site j without the own price variable, $\boldsymbol{\beta}_{-p,j}$ is the corresponding parameter vector, and β_{pj} is the coefficient on the own price variable. Given posterior simulations of $\boldsymbol{\alpha}_{(r)}^s$ as outlined in Step 1 above, draws of the choke prices under scenario s are obtained using equation (48). The consumer surplus for site j under conditions s in turn becomes:[15]

$$S_{fj}^s(\boldsymbol{\Gamma}, \boldsymbol{W}_f) = \sum_{k=1}^{K_j} max\left[c_{fj}^{k,s}(\boldsymbol{\Gamma}_{(r)}, \boldsymbol{W}_f) - P_{ij}^s, 0\right]. \tag{49}$$

[15] As with the case of forecasting trip changes, we set $T_{fj}^s = K_j - 1$ and as such will provide only a lower bound on consumer surplus as the last bin provides an aggregation of trips over $K_j - 1$.

A simulation-based estimate of $E[S^s_{fj}|\boldsymbol{q}^s_j, \boldsymbol{X}^s_f, \boldsymbol{Z}^s_f]$ follows naturally as:

$$\hat{E}[S^s_{fj}|\boldsymbol{q}^s_j, \boldsymbol{X}^s_f, \boldsymbol{Z}^s_f] = \frac{1}{R}\sum_{r=1}^{R} S^s_{fj}(\boldsymbol{\Gamma}_{(r)}, \boldsymbol{W}_f). \tag{50}$$

7 An Application

The data set we use to illustrate our proposed methodology is drawn from the Iowa Lakes Project, a four year study of recreational lake usage in the state. Funded by both the Iowa Department of Natural Resources and the U.S. Environmental Protection Agency, the goal of the project is to better understand visitation patterns to 132 of the state's key recreational lakes and to determine how these patterns are influenced by water quality. Understanding the linkage between water quality and recreational lake activity is important to evaluating and prioritizing efforts to comply with the Clean Water Act and to restore water quality in the state. Currently, roughly half of Iowa's lakes are currently considered as impaired by the National Water Quality Inventory (USEPA, 2000) .

Iowa is particularly well suited to studying the linkage between water quality and recreational activity, in part because of its wide range of water quality. Its lakes are among some of the best and worst in the nation in terms of water quality. For example, Secchi transparency, which roughly corresponds to the lake depth at which the bottom of a lake can still be seen, ranges from 0.09 meters (or 3.5 inches) to 5.67 meters (or 18.6 feet). Moreover, the lakes are heavily used by residents. In 2002, more than sixty percent of Iowa households visited at least one of the 132 key lakes in the state, with an average number of visits per household in excess of eight.

The primary source of data used in our application is the 2002 Iowa Lakes Survey. In November of that year, 8000 Iowa households were asked to complete a mail survey detailing their visits to each of the 132 key lakes in the state.[16] Standard Dillman (1978) procedures were used to ensure a high response rate, including follow-up reminders and a monetary incentive. Of the 8000 surveys mailed, 4423 surveys were returned, providing for an overall response rate of sixty-two percent once non-deliverables were accounted for. In addition to providing information about lake usage, standard socio-demographic variables were also collected. The first three rows of Table 2 provide summary statistics regarding trips taken by the 3859 households used in our analysis,[17] along with information regarding household income and the gender of the survey respondents. The latter two variables constitute the individual-specific factors included in the vector $\boldsymbol{z}_i$ of our model.

[16] See Azevedo *et al.* (2003) for additional details regarding the survey design and administration.

[17] Of the 4423 initial survey responses, we eliminated from our analysis those households who (a) did not complete the lake visitation portion of the survey, (b) were subsequently determined to live outside of Iowa (i.e., nonresidents), and (c) households reporting to have taken more than 52 trips to lakes in the state. The latter exclusion constitutes a small fraction of the sample reduction and reflects our intention to model days visits and exclude those "visits" by local residents who simply were passing by a lake.

The trip data provided by the 2002 Iowa Lakes Survey does not correspond directly to the ordinal response variables in our model (i.e., the y_{ij}'s). In particular, few households visit any individual lake more than five times, making the estimation of more than a small number of cutpoints (e.g., beyond four or five) difficult or impossible for all but a few of the sites. Moreover, most trip counts beyond ten are typically multiples of five, suggesting that the reported trips are rounded number recollections provided by the survey respondent, rather than a precise recollection of the trips taken. In our application, we aggregate trip counts in the upper tail of a site's frequency distribution into a single category, allowing K_j to vary by site. Table 3 summarizes the resulting frequency distribution of y_{ij} for each site, along with the mean and maximum trip values among households included in the last ordinal trip category.

Two additional sources of information are used in our analysis. First, the round trip travel cost from each survey respondent's residence to each of the twenty-nine lakes was calculated by first determining the associated round-trip travel distance and travel time using the software package *PCMiler (Streets Version 17)*. The out-of-pocket component of travel cost was computed as the round-trip travel distance multiplied by \$0.25 per mile.[18] To this figure was added the opportunity cost of time, calculated as one-third the estimated round-trip travel time multiplied by the respondent's average wage rate.[19] The fourth row Table 2 provides summary statistics for the resulting travel cost variable, labeled as $x_{ij(1)}$ in our model. While the model structure in Section 3 would allow for a full range of price and cross-price effects, for the sake of parsimony we assume that the travel costs to other sites have the same marginal effect (i.e., the same cross-price parameters) on travel to a given site. Thus, in addition to an own-price (travel cost) term, a second cross-price variable is included in our model as:

$$x_{ij(2)} \equiv \sum_{h \neq j} x_{ih(1)}. \tag{51}$$

Second, the hierarchical structure of the model in Section 5 allows the mean of the site-specific constant, α_j, to be influenced by site characteristics (i.e., the $\boldsymbol{q}_j$'s). In our application, we use Secchi transparency, one of the most commonly used limnological indicators of water quality, as our single site characteristic. Specifically, the Secchi transparency corresponds to the mean Secchi level measured by Iowa State University's Limnology Laboratory, averaged over three readings taken at each lake during the summer of 2002.

[18] The \$0.25 per mile is used as a relatively conservative estimate of gasoline and deprecation costs per mile of driving. This estimate is generally less than most official government reimbursement rates.

[19] The "average wage rate" is calculated for all respondents as their household's income divided by 2,000 (assuming total annual hours worked is 40 hours per week for 50 weeks).

8 Results

Using the model described in Section 3 and the posterior simulator detailed in Section 4, we fit the twenty-nine equation model using the Iowa Lakes data. We ran the posterior simulator 30,000 times, discarding the first 25,000 iterations as the burn-in. A single iteration in our simulator was available after approximately 35 seconds of computing time, so that 10,000 simulations were available after approximately 4 days of calculations. The relatively large burn-in was used as a precaution given what was found to be a slow mixing among the cutpoint simulations. We attribute this slow mixing to the rather sparsely-represented bins associated with many of the $y_{ij} \neq 0$ categories in our particular application; there is simply little information in the data that can be used to estimate many of the needed threshold parameters. In this regard we note that it may prove useful to elaborate the model in Section 3 to improve the mixing of the simulations when the data have many sparsely populated cells. Currently, the cutpoint values are treated as conditionally independent across sites. However, it is natural to think that the overall pattern of trip behavior may be similar across sites, and thus the corresponding elements δ_j may be related. Such a possibility could be incorporated through the adoption of a hierarchical prior for the δ_j which would allow the model to essentially borrow information from other sites when estimating the cutpoint parameters for a given site. We do not, however, explore this possibility in the current paper, though it is worth noting the potential benefits of such a specification, particularly when the data are decidedly unbalanced across the discrete outcomes.

8.1 Estimation Results

Table 4 provides the posterior means and $P(\cdot > 0|y)$ for each of the site specific parameters. Though fundamentally different in interpretation, these probabilities are similar in spirit to frequentist p-values. All of the conditioning variables (i.e., the $\boldsymbol{x}_{ij}$'s and $\boldsymbol{z}_i$'s) are normalized to have a zero mean to facilitate interpretation of the intercept terms. Gender is also scaled by a factor of ten.

The alternative specific constants, α_j, are uniformly negative for all of the sites, with over 99.9 percent of the posterior mass lying below zero. This is consistent with the fact that most individuals do not visit any given site, requiring the corresponding latent index y_{ij}^* to lie below zero. The own-price coefficients are likewise consistently negative, with the marginal impact of price on the latent index y_{ij}^* ranging from -0.73 for site 16 to -3.27 for site 20. In each case, the vast majority of posterior mass lies in the negative region. The cross price parameters are generally positive, as one might expect, with $P(\cdot|y) > 0.95$ for twenty of the twenty-nine sites. However, three sites (7, 13, and 26) have negative cross-price coefficients.

Turning to the individual household characteristics, income and gender (where gender=1 for males), we find that both variables tend to be positively relative to the propensity to visit a given site, though the impact is somewhat more consistent in the case of gender. The posterior means of the income coefficient are positive for twenty of the twenty-nine sites, with $P(\cdot|y) > 0.95$ in sixteen of these cases. While the remaining nine sites have posterior income parameter means that are negative, $P(\cdot|y) < 0.05$ in only three of these cases. The posterior mean of the gender coefficient is positive for all of the sites, a result consistent with the typical finding in the literature that males have a greater propensity to participate in outdoor recreation. For nineteen of the sites $P(\cdot|y) > 0.95$ for the male dummy, while for all but four of the sites $P(\cdot|y) > 0.90$. The cutpoint parameters are summarized in Table 5, though their direct interpretation is not of particular interest.

The hierarchical parameters, as described in Table 6, generally follow the patterns established by their site-specific counterparts. The intercept α_{01} is negative as expected, with $P(\alpha_{01} > 0|y)$ <0.001. Secchi transparency is positively correlated with a higher site specific constant, consistent with our expectations that sites with clearer water would, *ceteris paribus*, attract more visitors. However, this result should be interpreted with some caution, as $P(\alpha_{02} > 0|y) = 0.88$. The own- and cross-price coefficients are generally negative and positive, respectively, though the result for the cross-price term is less definitive, with $P(\beta_{02} > 0|y) = 0.76$. Finally, the hierarchical mean of gender coefficient is clearly positive, while the income coefficient γ_{01} is less clearly signed, reflecting heterogeneous impacts of income visitation to the various sites.

The final parameters of interest are the elements of the correlation matrix Σ. For the sake of space, the 406 elements of this matrix are not reported here. However, we note that the individual correlations vary substantially, with posterior means ranging from -0.443 to 0.831, suggesting that the flexibility allowed by our model in terms of the sign and size of the individual correlation elements is important in the current application. We also note that for more than half of the individual correlations (263 cases to be precise), the correlations are convincingly positive, with $P(\rho_{jj'} > 0|y) > 0.90$ for $j \neq j'$. In thirty-one cases, the correlations are convincingly negative, with $P(\rho_{jj'} > 0|y) < 0.10$ for $j \neq j'$.

8.2 Counterfactual Analysis

The results reported in the previous section provide a characterization of the demand for trips to each of the twenty-nine sites included in our recreation choice set. As described in section 6, these results can in turn be used to perform counterfactual analysis assessing the impact of changing individual and site characteristics, such as the cost of accessing the site and site quality attributes. As an illustration of this capability of our model we analyze changes in trip predictions, participation and consumer surplus measures for three hypothetical scenarios defined as follows:

- **Scenario 1: The Loss of Saylorville Reservoir**. We consider the loss of the most frequently visited lake in our choice set set (Saylorville Reservoir), which is lake twenty-one among the available alternatives;

- **Scenario 2: Increased Gasoline Prices**. We consider an increase in the out of pocket travel cost of driving to each site from $0.25 to $0.35 per mile to reflect recent increases in gas prices. This increase corresponds roughly to the change in the average price of gasoline from 2002 to 2007. Because the overall travel cost is composed of both out-of-pocket cost and the individual's opportunity cost of time, the change in gasoline prices will disproportionately impact lower income households;

- **Scenario 3: Improved Water Quality**. We consider an improvement to the water quality of lakes in the state. Specifically, we evaluate an improvement in the Secchi Transparency of each lake to a minimum level of 1.3 meters (or 4.3 feet), the median Secchi reading of the nonimpaired lakes in Iowa. This change impacts twenty-one of the twenty-nine lakes in our choice set and is consistent with state government objectives of removing all of the lakes from the EPA's list of impaired waterways.

These scenarios are designed to illustrate the advantages of our system approach to count data modeling, with attention given to the importance of the availability of substitutes. We expect the decrease in total trips made after a site is eliminated to be partially offset by increased trips to sites found to be substitutes, with the corresponding consumer surplus loss partially mitigated by increased value flowing from remaining substitute sites. Likewise, the behavioral and welfare effects of an increase in gas prices may be alleviated if the spatial configuration of lakes relative to the person's home provides reasonable substitutes closer to home. Finally, the value of improvements in water quality may hinge on the degree to which improved lakes have have substitutes in close proximity. All of these are context dependent, empirical issues that require a fully characterized demand system to address.

The results for our counterfactual experiments are shown in Table 7. The figures reported reflect totals for all 29 sites in the choice set. That is, the change in expected trips reflects total trips to all sites while the change in consumer surplus should be interpreted as the expected change in total surplus from all the sites during the recreation season. Likewise, participation is defined as an agent taking at least one trip to at least one of the available sites. The participation probability is the probability that a person switches participation status in response to an exogenous change. For improvements in site conditions this is the probability a person is a non-participant at baseline conditions and a participant under changed conditions. For degraded site conditions it is the opposite: the probability a person is a participant under baseline and a non-participant under changed conditions.

Several insights emerge from our counterfactual analysis. First, eliminating the most frequently visited site (Saylorville Reservoir) causes an average decrease of just over one-third of a trip. This relatively small effect is close to the sample average for trips taken to Saylorville Reservoir, suggesting there is little substitution to alternative sites for this particular scenario. Indeed, we find that agents on average increase their visits to the other, remaining sites by only 0.008 trips, though the data provide strong evidence of *some* increase in trips taken to these sites with $(P(\cdot > 0|y) = 1.00)$. The consumer surplus loss of around \$13 per person therefore reflects little additional value afforded by the remaining sites as viable substitutes. Since we have calculated the posterior mean baseline consumer surplus from all sites to be approximately \$77, the loss of this site constitutes a decrease in value of perhaps 17 percent. In addition we find a posterior mean probability of nearly 0.03 that a person will be a participant prior to the site loss and a non-participant after the site loss.

A different story emerges when we look at the increase in gas prices. Here the scenario is less extreme than losing a site, but the impact is spatially widespread in that all sites are more expensive. We find that people reduce their overall visits on average by one-third of a trip and suffer a loss of \$8.48 in consumer surplus per season. This is a smaller effect compared to the site loss. However, we find that a larger number of people are likely to become non-participants due to this scenario. In particular, we find there is a 0.06 probability that a person will visit at least one lake prior to the price increase, but quit taking visits when gas prices increase.

Finally, our quality improvement scenario does not lead to a large increase in trip taking (only 0.22 average trips per person) but does provide an increase in consumer surplus of \$8.67 per person on average (11 percent increase in total resource value). In addition we estimate a 0.03 probability that a person who does not visit under baseline conditions will visit when quality improves. Each of these estimates are less precise than their counterparts discussed above, owing to greater posterior uncertainty associated with the coefficient on Secchi transparency.

9 Discussion and Conclusions

We have described a Bayesian posterior simulator for flexibly fitting a high dimensional system of ordinal outcome equations. Since our emphasis is on examining how the approach may be useful for modeling correlated count outcomes we have labeled our approach a Generalized System of Count Outcomes (GeSCO) model, and examined its performance in the context of demand system analysis when realizations are non-negative integers. With our application to recreation demand we have demonstrated the model's flexibility in gauging how price and non-price attributes of the recreation sites influence behavior and how policy-induced changes in these attributes may lead to changes in peoples' visitation patterns and well-being.

We draw three conclusions from our modeling effort and applied exercise. First, our proposed methodology, similar to that of Chib *et al.* (2007), has genuine promise as a tool for analyzing the multivariate, correlated count outcomes that can occur in many areas of applied analysis. Our latent variables approach provides a flexible accounting for excess zeros and cross-equation correlation that cannot be easily reproduced by common alternatives such as Poisson or negative binomial systems. Furthermore, our posterior simulator is straightforward to implement relative to models offering comparable generality, and simulated data experiments imply it performs well in practice. These characteristics suggest additional evaluations of our approach in a wider range of contexts would be valuable extensions. The expanding availability of micro, survey, and confidential data suggests an increase in applied work using multivariate count methods such as we have proposed. Indeed, applications of our approach are likely to be found in areas of demand analysis that include empirical IO, transportation, and marketing as well as other areas of environmental economics.

Second, our proposed techniques for conducting posterior predictions of counterfactual scenarios allowed us to conduct welfare analysis in a way that represents an improvement over current practice in at least two dimensions. Most importantly, unobserved heterogeneity and the integer nature of the data are accommodated in posterior inference. This is in contrast to most studies employing traditional count distributions, which use the expected demand specification for prediction and consumer surplus calculations. Since the expected demand equation is continuous, an important aspect of the data generating process is ignored in counterfactual analysis. Our model allows computation of consumer surplus changes from step-wise demand functions that are simulated in a way that accounts for person and site-specific heterogeneity. Also, our specification includes a richer accounting for income, cross-price, and site specific heterogeneity effects than is typical in count data demand modeling. The latter is perhaps most important. The hierarchical structure we use to model site-specific intercepts bears similarity to work by Berry et al. (1995) and Bayer and Timmins (2007) from a classical perspective, in which the objective is to measure the effects of measurable attributes of the choice alternatives while explicitly allowing unobserved attributes to play a role. Our Bayesian solution to this problem is similar to that discussed by Yang et al. (2003), but generalized for use beyond the discrete choice context previously considered.

Finally, there are extensions and additional research that can be pursued within the context of our GeSCO framework. As discussed above some data environments (i.e. sparseness of observations in some of the discrete bins) may lead to poor precision and mixing in the posterior simulations for the cutpoints in the model. Additions to the modeling structure, such as hierarchical priors relating the cutpoints across the equations, may add in the identification of the cutpoints for some applications. A more conceptual extension would be to consider how restrictions from demand theory can be imposed on the structure of the problem in a way that maintains the flexibility of our approach and enables exact welfare analysis rather than the approximation approach pursued in this paper.

References

[1] Aitchison, J. and C. Ho (1989). The multivariate Poisson lognormal distribution, *Biometrica*, 76, 643-653.

[2] Albert, J. and S. Chib (1993). Bayesian analysis of binary and polychotomous response data, *Journal of the American Statistical Association*, 88, 669-679.

[3] Azevedo, C., K. Egan, J. Herriges, and C. Kling (2003). *Iowa Lakes Valuation Project: Summary and Findings from Year One.* Final Report to the Iowa Department of Natural Resources, August.

[4] Bayer, P., and C. Timmins (2007). Estimating equilibrium models of sorting across locations, *Economic Journal*, 117, 353-374.

[5] Berry, S., J. Levihsohn, and A. Pakes (1995). Automobile prices in equilibrium, *Econometrica*, 63, 841-890.

[6] Cameron, A. and P. Trivedi (1998). *Regression analysis of count data*, Cambridge, Cambridge University Press.

[7] Chen, M-H. and D. Dey (2000). Bayesian analysis for Correlated Ordinal Data Models, in *Generalized Linear Models: A Bayesian Perspective* D. Dey, S. Ghosh and B. Mallick, eds., 133-157. New York: Marcel-Dekker.

[8] Chib, S., J. Graves, I. Jeliazkov and M. Kutzbach (2007). Fitting and Comparison of Models for Multivariate Ordinal Outcomes, *Advances in Econometrics, Volume 23*, forthcoming.

[9] Chib, S., E. Greenberg, and R. Winkelmann (1998). Posterior simulation and Bayes factors in panel count data models, *Journal of Econometrics*, 86, 33-54.

[10] Chib, S. and R. Winkelmann (2001). Markov chain Monte Carlo analysis of correlated count data, *Journal of Business and Economic Statistics*, 19, 428-435.

[11] Cowles, M. (1996). Accelerating monte carlo markov chain convergence for cumulative-link generalized linear models, *Statistics and Computing*, 6, 101-111.

[12] Dillman, D. A. (1978) *Mail and Telephone Surveys – The Total Design Method*, New York: Wiley.

[13] Dionne, G., R. Gagne, F. Gagnon, and C. Vanasse (1997). Debt, moral hazard and airline safety: an empirical analysis, *Journal of Econometrics*, 79, 379-402.

[14] Egan, K. and J. Herriges (2006). Multivariate count data regression models with individual panel data from an on-site sample, *Journal of Environmental Economics and Management*, 52, 567-581.

[15] Englin, J. and J. Shonkwiler (1995). Estimating social welfare in count models, *Review of Economics and Statistics*, 77, 104-112.

[16] Englin, J., P. Boxall, and D. Watson (1998). Modeling recreation demand in a Poisson system of equations: an analysis of the impact of exchange rates, *American Journal of Agricultural Economics*, 80, 255-263.

[17] Geweke, J. (1991). Efficient simulation from the multivariate Normal and Student-t distributions subject to linear constraints, in: *Computer Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface* (ed. E. Keramidas), Interface Foundation of North America, Inc., Fairfax, 571-578.

[18] Hausman, J (1981). Exact consumer surplus and deadweight loss, *American Economic Review* 74(4), 662-76.

[19] Hellström, J (2006). A bivariate count data model for household tourism demand, *Journal of Applied Econometrics* 21, 213-226.

[20] Herriges, J., and D. Phaneuf (2002). Inducing Patterns Correlation and Substitution in Repeated Logit Model of Recreation Demand, *American Journal of Agricultural Economics*, 84, 1076-1090.

[21] King, G. (1989). A seemingly unrelated Poisson regression model, *Sociological Methods and Research*, 17, 235-255.

[22] Koop, G., D.J. Poirier and J.L. Tobias (2007). *Bayesian Econometric Methods*, Cambridge, Cambridge University Press.

[23] LaFrance, J. (1992). Incomplete Demand Systems, Weak Separability, and Weak Complementarity. Tucson: University of Arizona, Department of Agricultural and Resource Economics, Working Paper #77, December.

[24] LaFrance, J., and M. Hanemann (1989). The dual structure of incomplete demand systems, *American Journal of Agricultural Economics*, 719, 262-274. .

[25] Li, M. and J. L. Tobias (2005). Bayesian analysis of structural effects in an ordered equation system, *Studies in nonlinear dynamics and econometrics*, forthcoming.

[26] Lindley, D. and A. F. M. Smith (1972). Bayes estimates for the linear model, *Journal of the Royal Statistical Society*, Series B, 34, 1-41.

[27] Mullahy, J. (1997). Instrumental variable estimation of count data models: applications to models of cigarette smoking behavior, *Review of Economics and Statistics*, 79, 586-593.

[28] Nandram, B. and M.-H. Chen (1996). Reparameterizing the generalized linear model to accelerate Gibbs sampler convergence, *Journal of Statistical Computation and Simulation*, 54, 129-144.

[29] Ozuna, T. and I. Gomez (1994). Estimating a system of recreation demand functions using a seemingly unrelated Poisson regression approach, *Review of Economics and Statistics*, 76, 356-360.

[30] Ruser, J. (1991). Workers' compensation and occupational injuries and illnesses, *Journal of Labor Economics*, 9, 325-350.

[31] Train, Kenneth (2003). *Discrete Choice Methods with Simulation*, Cambridge, Cambridge University Press.

[32] U.S. Environmental Protection Agency (2000). "Nutrient Criteria Technical Guidance Manual: Lakes and Reservoirs." Office of Water, Office of Science and Technology, Report EPA-822-B00-001, Washington, D.C.

[33] von Haefen, R. and D. Phaneuf (2003). Estimating preferences for outdoor recreation: a comparison of continuous and count data demand system frameworks, *Journal of Environmental Economics and Management*, 45, 612-630.

[34] Wang, P., I. Cockburn, and M. Puterman (1998). Analysis of patent data: a mixed Poisson regression model approach, *Journal of Business and Economic Statistics*, 16, 27-41.

[35] Winkelmann, R (2000). Seemingly unrelated negative binomial regression, *Oxford Bulletin of Economics and Statistics*, 62, 553-560.

[36] Winkelmann, R. (2004). Health care reform and the number of doctor visits - an econometric analysis, *Journal of Applied Econometrics*, 19, 455-472.

[37] Yang, S., Y. Chen, and G. Allenby (2003). Bayesian analysis of simultaneous demand and supply, *Quantitative Marketing and Economics*, 1, 251-275.

Table 1: Posterior Means and True Values of Parameters in Generated Data Experiment

Parameter	Alternative											
	$j=1$		$j=2$		$j=3$		$j=4$		$j=5$		Common	
	Mean	True	Mean	True	Mean	True	Mean	True	Mean	True	Mean	True
α_j	.45	.40	.39	.36	-.05	-.02	.51	.46	.29	.19		
$\beta_j(1)$	-.55	-.47	-.43	-.43	-.56	-.57	-.58	-.57	-.48	-.45		
$\beta_j(2)$	.36	.43	.21	.24	.31	.30	.40	.39	.38	.46		
$\gamma_j(1)$	.11	.12	.01	.02	.01	.01	-.16	-.12	-.02	-.05		
$\gamma_j(2)$	.30	.28	.13	.13	.06	.02	.06	.07	.14	.16		
$\delta_2^{(j)}$	1.26	1.2	.51	.50	.46	.50	.30	.30	.27	.25		
$\delta_3^{(j)}$			.96	1.0	.94	1.0	.92	.90	.59	.55		
$\delta_4^{(j)}$			1.59	1.60	1.39	1.50			1.03	1.00		
$\delta_5^{(j)}$									1.36	1.35		
$\alpha_0(1)$											.43	.40
$\alpha_0(2)$											.23	.30
$\beta_0(1)$											-.50	-.50
$\beta_0(2)$											.32	.30
$\gamma_0(1)$											-.02	.00
$\gamma_0(2)$											.14	.10

Table 2: Summary Statistics

Variable	Model Variable	Mean	Std. Dev.	Min	Max
Total Day Trips (2002)	$T_{i\cdot}$	3.63	7.13	0	52
Household Income ($1000s)	$z_{i(1)}$	56.0	37.2	7.5	200.0
Gender (Male=1, Female=0)	$z_{i(2)}$	0.68	0.46	0	1
Travel Cost ($100's)	$x_{ij(1)}$	1.47	0.80	0.01	6.81
Cross Price ($100's)	$x_{ij(2)}$	41.11	15.98	17.31	132.87
Secchi Transparency (m)	$q_{j(2)}$	1.15	1.12	0.09	5.67

Table 3: Ordinal Response Data

	Sample frequencies for y_{ij} =								Tail Category (K_j)		
Site	0	1	2	3	4	5	6	7	K_j	Mean(T_{ij})	Max(T_{ij})
1	91.9	2.9	1.6	1.0	0.5	0.8	1.3		7	13.0	25
2	95.7	2.0	1.0	1.4					4	5.9	15
3	98.2	0.7	1.1						3	6.8	30
4	98.7	0.5	0.8						3	4.4	25
5	97.2	1.0	0.7	1.1					4	5.3	15
6	92.6	3.1	1.4	0.8	2.2				5	10.9	50
7	91.6	3.4	1.6	0.8	0.7	2.1			6	11.3	50
8	97.8	1.1	0.5	0.5					4	6.7	20
9	94.1	2.6	0.9	0.7	1.7				5	7.0	20
10	98.9	0.6	0.5						3	10.0	35
11	98.2	0.7	1.0						3	4.6	20
12	97.5	0.8	0.6	1.2					4	10.3	50
13	93.5	2.5	1.3	0.6	0.7	1.3			6	8.3	20
14	98.5	0.7	0.9						3	3.9	15
15	99.2	0.4	0.4						3	3.6	10
16	98.2	0.9	0.9						3	3.1	15
17	98.5	0.7	0.8						3	3.7	20
18	95.5	2.0	1.0	1.6					4	8.2	45
19	93.0	3.0	1.5	0.5	2.0				5	10.5	40
20	98.2	0.8	1.0						3	4.9	25
21	87.8	4.2	2.6	1.2	0.7	0.8	0.8	1.9	8	13.4	50
22	99.2	0.3	0.5						3	6.2	15
23	99.4	0.3	0.3						3	3.8	15
24	96.7	1.5	0.5	1.2					4	8.0	50
25	98.0	0.8	0.5	0.7					4	5.0	10
26	99.7	0.2	0.1						3	7.2	20
27	99.7	0.2	0.1						3	8.5	24
28	98.5	0.7	0.9						3	3.2	6
29	92.9	3.0	1.4	0.7	2.1				5	7.9	30

Table 4: Posterior Means of Site-Specific Parameters

Site	Site									
	α_j		β_{1j} (Own Price)		β_{2j} (Cross Price)		γ_{1j} (Income)		γ_{2j} (Gender)	
	Mean	$P(\cdot > 0\|y)$	Mean	$P(\cdot > 0\|y)$	Mean	$P(\cdot > 0\|y)$	Mean	$P(\cdot > 0\|y)$	Mean	$P(\cdot > 0\|y)$
1	-2.95	0.0000	-2.10	0.0000	0.03	1.0000	0.01	0.7418	0.92	0.9458
2	-1.59	0.0000	-0.98	0.0000	0.02	1.0000	0.09	1.0000	1.64	0.9952
3	-3.41	0.0000	-2.48	0.0000	0.02	0.9728	0.09	1.0000	1.84	0.9680
4	-3.69	0.0000	-2.42	0.0000	0.06	1.0000	-0.11	0.0000	1.47	0.9758
5	-3.60	0.0000	-3.05	0.0000	0.06	1.0000	0.01	0.6408	1.73	0.9956
6	-2.32	0.0000	-1.98	0.0000	0.02	1.0000	0.10	1.0000	0.99	0.9360
7	-2.70	0.0000	-1.80	0.0000	-0.01	0.0232	0.10	1.0000	0.62	0.8566
8	-2.36	0.0000	-1.39	0.0000	0.04	1.0000	-0.02	0.1700	0.96	0.9158
9	-1.48	0.0000	-0.96	0.0000	0.02	1.0000	0.08	1.0000	1.10	0.9544
10	-2.74	0.0000	-1.32	0.0000	0.01	0.8282	0.06	1.0000	1.73	0.9988
11	-2.77	0.0000	-2.03	0.0000	0.06	1.0000	-0.04	0.0498	2.19	0.9996
12	-2.97	0.0000	-2.10	0.0000	0.01	0.8876	0.07	0.9994	1.10	0.9548
13	-2.96	0.0000	-1.94	0.0000	-0.01	0.0316	0.11	1.0000	0.79	0.8950
14	-2.58	0.0000	-1.19	0.0000	0.02	1.0000	0.03	0.9454	1.96	1.0000
15	-3.12	0.0000	-1.61	0.0000	0.02	1.0000	-0.02	0.1580	1.26	0.9428
16	-1.92	0.0000	-0.73	0.0000	0.01	0.9746	0.04	0.9880	0.32	0.6864
17	-2.40	0.0000	-1.11	0.0000	0.01	0.7754	0.05	1.0000	0.40	0.7512
18	-2.25	0.0000	-1.53	0.0000	0.02	1.0000	0.08	1.0000	2.33	1.0000
19	-2.70	0.0000	-2.02	0.0000	0.01	1.0000	0.07	1.0000	1.41	0.9888
20	-4.77	0.0000	-3.27	0.0000	0.03	1.0000	-0.02	0.1424	1.65	0.9948
21	-2.53	0.0000	-1.73	0.0000	0.00	0.5970	0.06	1.0000	0.88	0.9404
22	-2.34	0.0000	-1.24	0.0000	0.07	1.0000	-0.12	0.0000	2.28	1.0000
23	-2.54	0.0000	-1.09	0.0000	0.05	1.0000	-0.03	0.1616	1.50	0.9612
24	-2.19	0.0000	-1.15	0.0000	0.01	0.7776	0.06	0.9996	2.30	0.9992
25	-2.47	0.0000	-1.10	0.0000	0.02	1.0000	-0.01	0.3340	1.89	1.0000
26	-3.48	0.0000	-1.28	0.0000	-0.03	0.0006	0.13	1.0000	1.28	0.9568
27	-2.97	0.0000	-1.38	0.0000	0.02	0.9450	0.05	0.8884	1.36	0.9244
28	-2.48	0.0000	-1.00	0.0000	0.03	1.0000	-0.03	0.0662	2.85	1.0000
29	-1.37	0.0000	-1.06	0.0000	0.02	1.0000	0.10	1.0000	1.21	0.9594

Table 5: Posterior Means of Cutpoints

	Site					
Site	δ_2	δ_3	δ_4	δ_5	δ_6	δ_7
1	0.31	0.54	0.73	0.87	1.11	
2	0.33	0.63				
3	0.58					
4	0.75					
5	0.34	0.76				
6	0.37	0.63	0.82			
7	0.34	0.57	0.70	0.84		
8	0.38	0.74				
9	0.28	0.42	0.56			
10	0.56					
11	0.43					
12	0.32	0.68				
13	0.33	0.57	0.72	0.95		
14	0.48					
15	0.60					
16	0.48					
17	0.47					
18	0.30	0.53				
19	0.36	0.63	0.77			
20	0.65					
21	0.32	0.59	0.76	0.87	1.01	1.19
22	0.62					
23	0.63					
24	0.28	0.44				
25	0.26	0.54				
26	0.96					
27	0.96					
28	0.42					
29	0.31	0.52	0.66			

Table 6: Hierarchical Parameters

Parameter	Mean	$P(\cdot > 0\|y)$
α_{01}	-2.69	0.0000
α_{02} (Secchi)	0.10	0.84
β_{01} (Own-Price)	-1.60	0.0000
β_{02} (Cross-Price)	0.021	0.7340
γ_{01} (Income)	0.04	0.8312
γ_{02} (Gender)	1.41	1.0000

Table 7: Counterfactual Analysis

	Change in:		
Scenario	Expected Trips $E(T_{fj}^1 - T_{fj}^0)$	Participation $\hat{\Xi}$	Consumer Surplus $E(S_{fj}^1 - S_{fj}^0)$
1 (Site Loss)	-0.38	0.028	-$13.05
	(0.00)[a]	(1.00)	(0.00)
2 (Increased gasoline prices)	-0.33	0.058	-$8.48
	(0.00)	(1.00)	(0.00)
3 (Improved water quality)	0.22	0.028	$8.67
	(0.93)	(0.93)	(0.93)

[a] The values in parentheses are $P(\cdot > 0|\boldsymbol{y})$.

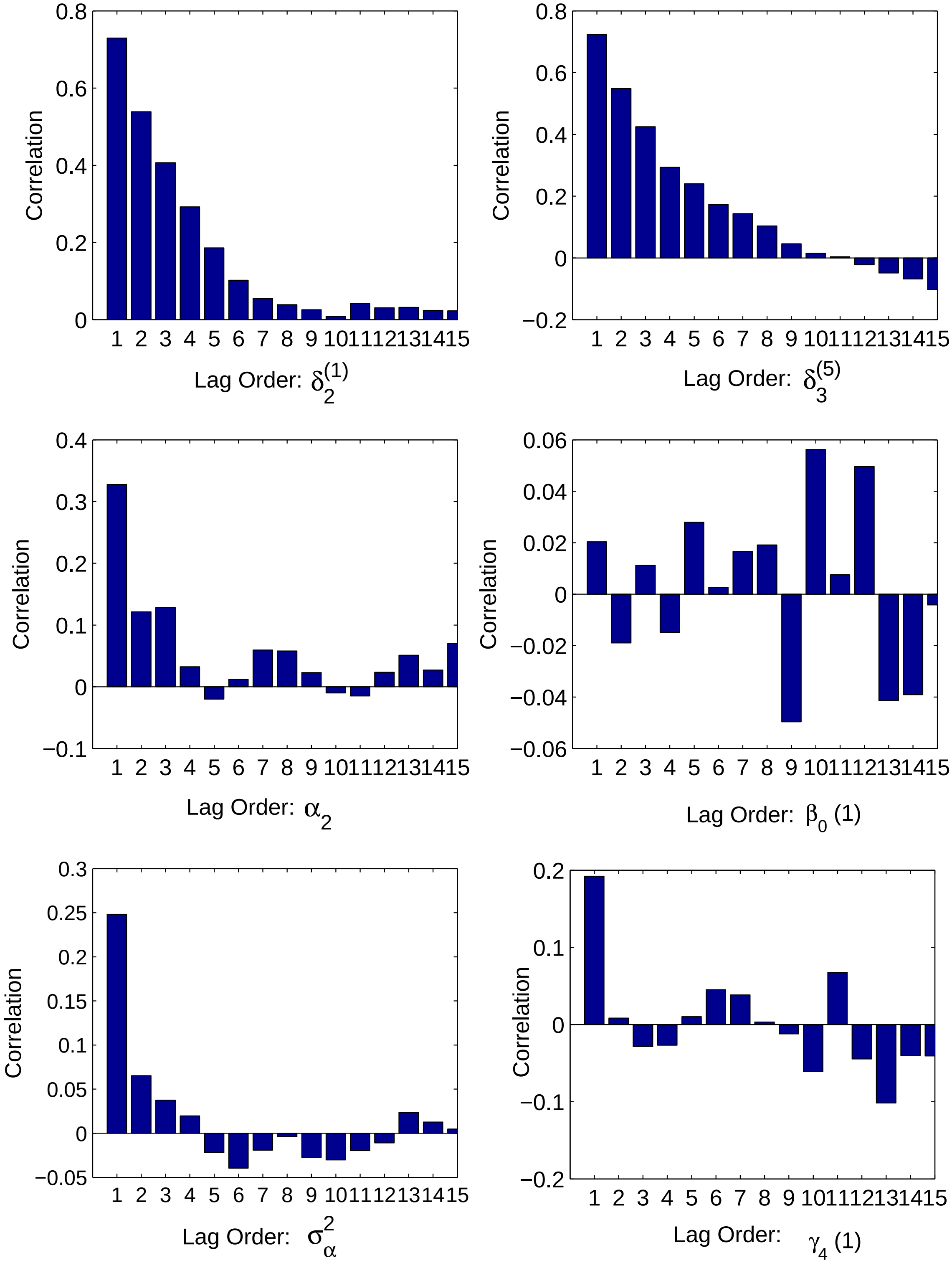

Figure 1: Lagged Autocorrelation for Select Parameters

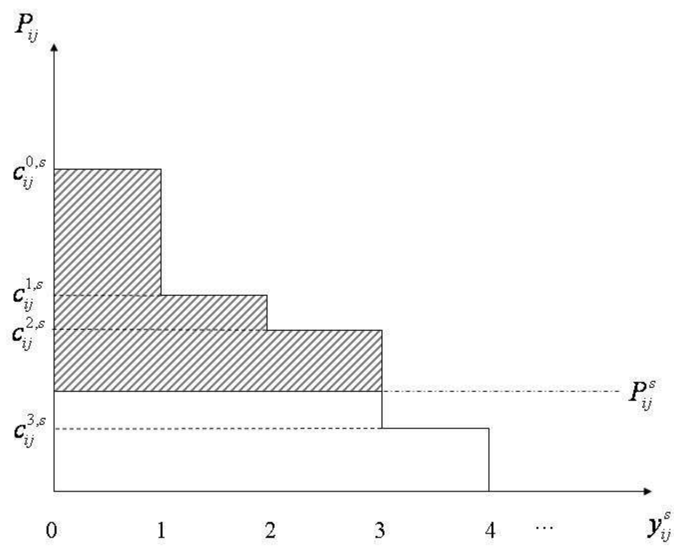

Figure 2: Consumer Surplus