Federal Reserve Bank of Minneapolis
Research Department Staff Report #170

May 1994

Alternative Computational Approaches to Inference in the Multinomial Probit Model

John Geweke, Michael Keane, and David Runkle*

Federal Reserve Bank of Minneapolis and
University of Minnesota

ABSTRACT

This research compares several approaches to inference in the multinomial probit model, based on Monte-Carlo results for a seven choice model. The experiment compares the simulated maximum likelihood estimator using the GHK recursive probability simulator, the method of simulated moments estimator using the GHK recursive simulator and kernel-smoothed frequency simulators, and posterior means using a Gibbs sampling-data augmentation algorithm. Each estimator is applied in nine different models, which have from 1 to 40 free parameters. The performance of all estimators is found to be satisfactory. However, the results indicate that the method of simulated moments estimator with the kernel-smoothed frequency simulator does not perform quite as well as the other three methods. Among those three, the Gibbs sampling-data augmentation algorithm appears to have a slight overall edge, with the relative performance of MSM and SML based on the GHK simulator difficult to determine.

*Geweke acknowledges partial financial support from NSF Grant SES-9210070. The authors acknowledge helpful discussions with Peter Rossi and Paul Ruud, and research assistance from Donna Boswell and Zhenyu Wang. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

1. Introduction

The multinomial probit is an appealing model of choice behavior because it allows a flexible pattern of conditional covariance among the latent utilities of alternatives. Nevertheless, multinomial probit applications have been limited because the required integrations of the multivariate normal density over subsets of Euclidean space are computationally burdensome. The computational simplicity of the multinomial logit has made it the model of choice for applied work. However, because the multinomial probit model relaxes the assumption of independence of irrelevant alternatives, it is generally preferred in principle to the multinomial logit model (McFadden 1984, pp. 1395–458). Recently the method of simulated moments (McFadden 1989, Pakes and Pollard 1989) and Gibbs sampling with data augmentation (Albert and Chib 1992, McCulloch and Rossi 1992) have shown promise of making the required computations in the multinomial probit model practical. The development of the highly accurate GHK probability simulator (see Geweke 1991, Hajivassiliou and McFadden 1990, and Keane 1990) has also led to renewed interest in simulated maximum likelihood (Albright, Lerman, and Manski 1977) as a method for estimating multinomial probit models.

The objective of the research reported here is to provide a systematic comparison of the numerical properties of different simulation based methods of inference in the multinomial probit model. Rather than considering the performance of these methods on a single model for a single data set, we attempt to control for a number of features of the inference problem, such as the number and nature of the unknown parameters of interest, and the information content of the data on which inference is based. Also, we investigate for the first time how the performance of MSM estimation is affected by the type of probability simulator employed (i.e., GHK vs. kernel smoothing).

While some investigators have examined the performance of *particular* estimators and computational techniques, and Hajivassiliou, McFadden, and Ruud (1992), Borsch-Supan and Hajivassiliou (1993), and Hajivassiliou (1992) have made systematic comparison of alternative probability simulators, we are aware of only one systematic comparison of different estimators *per se:* Keane (1994) compares method of simulated moments and simulated maximum likelihood estimators for an eight period binomial probit model in a Monte Carlo study. This paper is the first to compare performance of alternative methods of inference for the multinomial probit model, and the first to examine how the relative performance of alternative methods differs across model specifications and across different data sets.

In addition to addressing this main objective, this work introduces a new factor structure for the disturbances that may help to alleviate the proliferation of covariance matrix parameter problem in MNP models. We also illustrate Bayesian inference in a multinomial probit model with a factor structure for the first time. (See Elrod and Keane 1994 for a discussion of factor structures for probit models.)

The paper contains two separate Monte-Carlo experiments. In the first, we generate 10 artificial data sets from a seven alternative model, and compare the performance of the simulated maximum likelihood estimator using the GHK simulator, the MSM estimator using the GHK and kernel-smoothed frequency simulators, and posterior means using a Gibbs sampling-data augmentation algorithm. We consider 9 different model specifications in which restrictions are placed on different groups of parameters in order to determine how such restrictions affect the relative performance of the methods. In this experiment, the performance of the MSM and SML estimators using the GHK simulator, and the performance of the Bayesian inference procedure using the Gibbs sampling-data augmentation algorithm, all clearly dominate that of the MSM estimator based on kernel smoothing.

In the second experiment, we use actual data on ketchup purchases acquired from the Nielsen company to construct regressors and estimate parameter values for a seven alternative choice model. We then generate 50 artificial data sets using these parameter values and regressors. We then compare the performance of MSM and SML based on the GHK simulator, and posterior means using the Gibbs sampling-data augmentation algorithm. Here, significant biases appear for all three methods, suggesting that, given the configuration of data and parameter values, the sample size is not large enough for small sample bias to become negligible. Bayesian inference appears to have a slight edge in performance over the classical methods, with MSM-GHK and SML-GHK difficult to choose between. Given the computational cost of performing 50 replications, and given that MSM based on kernel smoothing was dominated by other methods in the first experiment, we do not include it in the second experiment.

The next section sets out the multinomial probit model and establishes notation, and Section 3 describes the design of the experiments. Section 4 discusses computational issues for classical inference, including probability simulation, the simulated maximum likelihood method, and the method of simulated moments. Section 5 does the same for Bayesian inference, based on the Gibbs sampler and data augmentation. The results of the experiments are presented in Section 6. Section 7 concludes by giving our overall assesement of the performance of the alternative methods across both experiments, presenting timing comparisons for the alternative methods, and discussing how computation times for the methods would be likely to differ in different modelling contexts.

2. The Probit Model

Let individual i choose among a set of mutually exclusive alternatives, $j = 1, \ldots, J$. Assume that i's utility from choice j is

$$(2.1)\quad u_{ij} = x_i'\beta_j + z_{ij}'\gamma_j + \epsilon_{ij}$$

where

x_i is a $k \times 1$ vector of individual characteristics (e.g., age, sex),

z_{ij} is a $p \times 1$ vector of alternative-specific attributes faced by individual i (e.g., price, quality),

ϵ_{ij} is the alternative-specific disturbance in i's utility from choice j.

The probit model is obtained by assuming the ϵ_{ij} have a multivariate normal distribution: $\epsilon_i = (\epsilon_{i1},\ldots,\epsilon_{iJ})' \sim \text{IIDN}(0,\Sigma)$.

The econometrician observes the utility-maximizing choice $\text{argmax}_j(u_{ij})$ made by each of N individuals, as well as the individuals' characteristics and the alternative-specific attributes they face, but does not observe the individual utilities. The econometrician desires to learn about the β_j, the γ_j, Σ, and the probability that individual i would make choice j.

The model (2.1) may be written in stacked form,

$$u_i = X_i\beta + Z_i\gamma + \epsilon_i,$$

where $u_i = (u_{i1},\ldots,u_{iJ})'$ and the arrangement of $X_i(J \times q)$, $Z_i(J \times r)$, $\beta(q \times 1)$, and $\gamma(r \times 1)$ reflect cross-equation as well as zero restrictions on the β_j and $\gamma_j (q \leq Jk,\ r \leq Jp)$. It is well known that this model is unidentified (see Bunch 1991 or Dansie 1985). Identification is achieved by working with the differenced system

(2.2) $u_i^* = X_i^*\beta^* + Z_i^*\gamma^* + \epsilon_i^*, \quad (i = 1,\ldots,N),\ \epsilon_i^* \sim \text{IIDN}(0,\Sigma^*)$

where u_i^* is a J × 1 vector with $u_{ij}^* = (u_{ij}-u_{iJ})/(\sigma_{11}-2\sigma_{1J}+\sigma_{JJ})^{1/2}$, $(j = 1,\ldots,J-1)$; $u_{iJ}^* = 0$; $\text{var}(\epsilon_{i1}^*) = 1$; and X_i^*, Z_i^*, β^*, and γ^* are the corresponding appropriate transformations of X_i, Z_i, β, and γ, respectively. (We provide a specific example of these transformations in the next section.)

The corresponding log-likelihood function is

(2.3) $$L_N(\beta^*,\gamma^*,\Sigma^*) = N^{-1} \sum_{i=1}^{N}\sum_{j=1}^{J} d_{ij} \ln P(j\,|\,\beta^*,\gamma^*,\Sigma^*,X_i^*,Z_i^*),$$

where $d_{ij} = 1$ if individual i chooses j and $d_{ij} = 0$ otherwise, and $P(j\,|\,\beta^*,\gamma^*,\Sigma^*,X_i^*,Z_i^*) = P(u_{ij}^* \geq u_{ik}^*\ \forall\ k\,|\,\beta^*,\gamma^*,\Sigma^*,X_i^*,Z_i^*)$. This requires the evaluation of a $(J-1)$-dimensional integral, which is computationally burdensome using classical methods (like quadrature) when $J > 4$ (Kahaner 1991).

3. Experimental Design

A. Experiment One

In the first experiment, the comparison of different methods of inference is made using artificial data generated from a seven alternative model,

$$u_{ij} = \beta_{j1} + \beta_{j2}x_i + \gamma_{j1}z_{ij} + \epsilon_{ij}, \quad (j = 1,\ldots,7);$$
$$\epsilon_i = (\epsilon_{i1},\ldots,\epsilon_{i7})' \sim \text{IIDN}(0,\Sigma).$$

Therefore, in the transformed model

(3.1) $u_{ij}^* = \beta_{j1}^* + \beta_{j2}^*x_i^* + \gamma_{j1}^*z_{ij}^* + \gamma_{71}z_{i7}^* + \epsilon_{ij}^*, \quad (j = 1,\ldots,6);\ u_{i7}^* = 0;$

(3.2) $\epsilon_i^* = (\epsilon_{i1}^*,\ldots,\epsilon_{i6}^*)' \sim \text{IIDN}(0,\Sigma^*).$

For our data generating process we adopt the parameter values $\beta_{j1} = \beta_{j2} = \gamma_{j1} = 1$ and $\Sigma = I$. In the transformed model, this gives:

(3.3) $x_i^* = x_i$, $z_{ij}^* = z_{ij}$, $\beta_{j1}^* = \beta_{j2}^* = 0$, $\gamma_{j1}^* = 1/\sqrt{2}$, $\gamma_{71} = -1/\sqrt{2}$, and $\Sigma^* = 0.5(I+ee')$,

where e denotes a 6×1 vector of units. The covariates were generated $(x_i,z_{i1},\ldots,z_{i7}) \sim IIDN(0,I_8)$. The experiment used 10 artificial data samples generated from this model, each with observations on 5,000 individuals.

Nine different models were considered. The nine models are the Cartesian product of three alternative specifications for the coefficients, and three alternative specifications for the variance matrix of the disturbances.

The first coefficient specification is as in (3.1), where all coefficients are unrestricted. In the second coefficient specification, the coefficients γ_{j1} are restricted to be equal,

(3.4) $u_{ij}^* = \beta_{j1}^* + \beta_{j2}^* x_i + \gamma_{11}^*(z_{ij}-z_{i7}) + \epsilon_{ij}^*$, $(j = 1,\ldots,6)$.

In the third coefficient specification, the coefficients β_{j1} are restricted to be equal, as are the β_{j2},

(3.5) $u_{ij}^* = \gamma_{j1}^*(z_{ij}-z_{i7}) + \epsilon_{ij}^*$, $(j = 1,\ldots,6)$.

The first variance specification is as in (3.2) where Σ^* is unrestricted. In classical estimation, we parameterize Σ^* by its Choleski decomposition $\Sigma^* = AA'$, where A is a lower triangular matrix with typical element a_{ij}; $a_{11} = 1$. For Bayesian inference we report the posterior means of the a_{ij}, for comparison with the classical estimates. In the second variance specification Σ is diagonal with diagonal element $\lambda + \xi_j$ where $\xi_J = 0$. In differenced form this gives

(3.6) $\Sigma^* = \lambda(I+ee') + D$,

where $D = \mathrm{diag}(\xi_1,\ldots,\xi_{J-1})$. To achieve identification, impose $\lambda = 1/2$ to obtain $0.5(I+ee') + D$. Then divide all elements of this matrix by $1 + \xi_1$ in order to impose $\sigma^*_{11} = 1$. This gives

$$\Sigma^* = 0.5(1+\xi_1)^{-1}(I+ee') + (1+\xi_1)^{-1}D.$$

In the final variance specification we impose $\Sigma = I$ and $\Sigma^* = 0.5(I+ee')$, which are the true values for the assumed data generating process.

B. Experiment Two

In the second experiment, we work only with the second coefficient specification (3.4), in which the γ_{j1} are restricted to be equal, combined with the unrestricted covariance specification (3.2). However, rather than using the parameter values $\beta_{j1} = \beta_{j2} = \gamma_{j1} = 1$ and $\Sigma = I$ that were used in the first experiment, we estimate these coefficient values using an actual data set. Specifically, we use Nielsen scanner panel data on ketchup purchases by 1,153 households with 5,353 total purchase occasions to estimate the parameters of model (3.4), (3.2) via MSM based on the GHK simulator.

The Nielsen ketchup data is described in detail in Keane and Elrod (1994). The alternatives are 32 ounce sizes of DelMonte, Hunts, and Heinz, along with 18, 28, 40, and 64 ounce sizes of Heinz. The household specific characteristic x_i that we use in estimation is household size. The estimated model actually has four brand specific attribute variables z_{ij} for $j = 1, 7$. These are price and dummies for the presence of three different types of displays. To arrive at a model identical in form to (3.4), in which there is only a single brand specific attribute, we construct a single alternative specific variable which is a linear combination of the price and the three display dummies, constructed in such a way that the coefficient γ^*_{11} on the constructed z variable is equal to the estimated price coefficient. We then construct 50 artificial data sets of 5,000 observations each, using as model parameters the brand intercepts, household size and price coefficients and covariance matrix elements estimated on the Nielsen data, and using as covariates the household sizes and constructed z variables from the first 5,000 purchase occasions in the Nielsen data. Note that, although the actual data is a panel, since we estimate a model that assumes no serial correlation in the unobservables, the artificial data we construct also has no serial correlation.

Our goal in the second experiment is to provide a more stringent test of the performance of the three preferred simulation based approaches to inference. In the first experiment, the covariates were orthogonal. In this experiment the covariates are correlated. For example, prices of different brands of ketchup will tend to move together due to competitive reactions, and also because five of the alternatives are from the same manufacturer. In addition, certain households will tend to shop at low price stores, while others shop at higher price stores. Such correlations among covariates may reduce the information content of any sample of given size, so that larger sample sizes may be necessary before small sample bias becomes negligible. An additional factor is that the error structure estimated from the Nielsen ketchup data is substantially more complex than that assumed in experiment one. There we assumed a simple one factor error structure in the differenced model, with all error variances equal. A notable feature of the actual data is that the estimated error variances differ widely across alternatives. Hajivassiliou, McFadden, and Ruud find that the accuracy of probability simulators deteriorates as error variances become more unequal, so that the covariance structure assumed in the second experiment may be expected to lead to a deterioration in the performance of the simulation based estimators relative to their performance in experiment one.

4. Classical Inference Using Probability Simulators

Classical methods of estimation are all based on the log-likelihood function (2.3). They approximate the $P(j\,|\,\beta^*,\gamma^*,\Sigma^*,X_i^*,Z_i^*)$ using a probability simulator, and then apply conventional procedures to solve moment conditions or maximize the log-likelihood function. We turn first to three probability simulators, and then to two estimation procedures.

4.1 Alternative Probability Simulators

Frequency Simulator

Lerman and Manski (1981) proposed using a frequency simulator to approximate the probabilities appearing in (2.3). A frequency simulator is constructed as follows:

1. For each of M replications ($\ell = 1,\ldots,M$), draw a $J - 1$ vector of independent standard normal random variables, η_i^ℓ.
2. Let $\epsilon_i^{*\ell} = A\eta_i^\ell$ and compute $u_i^{*\ell} = X_i^*\beta^* + Z_i^*\gamma^* + \epsilon_i^{*\ell}$ for all ℓ.
3. Let $I_{[j]}^{(\ell)} = \text{argmax}_{j=1,\ldots,J} u_{ij}^{*\ell}$ for all ℓ.
4. Construct $\hat{P}_F(j\,|\,\beta^*,\gamma^*,\Sigma^*,X_i^*,Z_i^*) = \frac{1}{M}\sum_{\ell=1}^{M} I_{[j]}^{(\ell)}$.

This simulator is fast, easy to compute, and unbiased. However, the simulated probabilities $\hat{P}_F$ are discontinuous in the parameters β^*, γ^*, and Σ^*, necessitating use of derivative-free algorithms to maximize the simulated log-likelihood function obtained when $\hat{P}_F$ is substituted into (2.3). There is also a positive probability that $\hat{P}_F = 0$, in which case (2.3) cannot be computed at all after substitution.

Kernel-Smoothed Frequency Simulator

To avoid the discontinuities in crude-frequency simulator, McFadden (1989) proposed the use of a kernel-smoothed frequency (KS) simulator. For kernel-smoothing parameter ρ and each of M replications ($\ell = 1,\ldots,M$), begin with Steps 1 and 2 of the crude frequency simulator. Then construct

$$\hat{P}_{KS}(j \mid \beta^*, \gamma^*, \Sigma^*, Z_i^*, X_i^*) = \frac{1}{M} \sum_{\ell=1}^{M} e^{u_{ij}^{\ell*}/\rho} \left[1 + \sum_{j=1}^{J-1} e^{u_{ij}^{\ell*}/\rho} \right]^{-1} .$$

This simulator shares the advantages and disadvantages of kernel-smoothers generally. The larger the value of ρ, the smaller the variance in the estimator $\hat{P}_{KS}$, but the greater the bias. The appropriate choice of ρ will diminish as the number of simulations M increases. While rates of decrease sufficient to eliminate asymptotic bias are known (see Sections 4.2 and 4.3), the determination of which ρ to use with a given M requires some costly experimentation as a practical matter.

The GHK Recursive Simulator

The GHK recursive simulator, due to Geweke (1991), Hajivassiliou and McFadden (1990), and Keane (1990, 1994), is based on the observation that the choice probabilities in the multinomial probit model may be written as a sequence of conditional probabilities that may be simulated recursively. This simulator is of particular interest, because in a rather exhaustive study of many alternative probability simulators Hajivassiliou, McFadden, and Ruud (1992) concluded that GHK was the most accurate and reliable method of all those considered.

Some additional notation is needed to describe this simulator. Write equation j of (2.2):

$$u_j^* = X_j^* \beta^* + Z_j^* \gamma^* + \epsilon_j^* .$$

Here X^*_j denotes row j of X* and Z^*_j denotes row of j of Z*. The i subscript is dropped in this section only for ease of notation. Define:

$$\tilde{u}^j_k = u^*_k - u^*_j \quad \text{for } k = 1, \ldots, J, \text{ and } \tilde{\epsilon}^j = \epsilon^*_k - \epsilon^*_j \quad \text{for } k = 1, J,$$

where it is understood that $\epsilon^*_J = 0$. Further define the $(J-1) \times 1$ vectors:

$$\tilde{u}^j = (\tilde{u}^j_1,\ldots,\tilde{u}^j_{j-1},\tilde{u}^j_{j+1},\ldots,\tilde{u}^j_J)' \text{ and } \tilde{\epsilon}^j = (\tilde{\epsilon}^j_1,\ldots,\tilde{\epsilon}^j_{j-1},\tilde{\epsilon}^j_{j+1},\ldots,\tilde{\epsilon}^j_J)'.$$

Recall that alternative j is chosen if $u^*_k - u^*_j \leq 0$ for $k = 1, \ldots, J$ which is equivalent to the condition that all the elements of $\tilde{u}_j$ are less than or equal to zero.

Since $\tilde{\epsilon}^j$ is a linear transformation of ϵ^* the distribution of $\tilde{\epsilon}^j$ is IIDN$(0,\tilde{\Sigma}^j)$ where $\tilde{\Sigma}^j$ is the corresponding appropriate transformation of Σ^*. Let $\tilde{A}^j$ be the unique lower triangular Cholesky factorization $\tilde{\Sigma}^j = \tilde{A}^j(\tilde{A}^j)'$. Then $\tilde{\epsilon}^j = \tilde{A}^j\eta$ where $\eta = (\eta_1,\ldots,\eta_{j-1},\eta_{j+1},\ldots,\eta_J)'$ is a $(J-1) \times 1$ vector of independent standard normal random variables.

Define $\tilde{u}^j_k(\eta^\ell_1,\ldots,\eta^\ell_p)$ as the value of $\tilde{u}^j_k$ when the random variables η_1 through η_p are fixed at the draw $(\eta^\ell_1,\ldots,\eta^\ell_p)$, where $p \leq k$. For $p = k$, $\tilde{u}^j_k(\eta^\ell_1,\ldots,\eta^\ell_p)$ is a number, while for $p < k$ it is a random variable. Then, the GHK simulator for the probability of alternative j is constructed as follows:

(1) Draw η^ℓ_1 such that $\tilde{u}^j_1(\eta^\ell_1) < 0$ for $\ell = 1, \ldots, M$

⋮

(j−1) Draw η^ℓ_{j-1} such that $\tilde{u}^j_{j-1}(\eta^\ell_1,\ldots,\eta^\ell_{j-1}) < 0$ for $\ell = 1, \ldots, M$

(j) Skip η_j

(j+1) Draw η^ℓ_{j+1} such that $\tilde{u}^j_{j+1}(\eta^\ell_1,\ldots,\eta^\ell_{j-1},\eta^\ell_{j+1}) < 0$ for $\ell = 1, \ldots, M$

⋮

(J−1) Draw η^ℓ_{J-1} such that $\tilde{u}^j_{J-1}(\eta^\ell_1,\ldots,\eta^\ell_{j-1},\eta^\ell_{j+1},\ldots,\eta^\ell_{J-1}) < 0$ for $\ell = 1, \ldots, M$

and finally, construct:

$$\hat{P}_{GHK}(j\,|\,\beta^*,\gamma^*,\Sigma^*,X^*,Z^*)$$

$$= \frac{1}{M}\sum_{\ell=1}^{M} P(\tilde{u}_1^j < 0) \prod_{k=2}^{j} P[\tilde{u}_k^j(\eta_1^\ell,\ldots,\eta_{k-1}^\ell) < 0] \prod_{k=j+1}^{J-1} P[\tilde{u}_k^j(\eta_1^\ell,\ldots,\eta_{j-1}^\ell,\eta_{j+1}^\ell,\ldots,\eta_k^\ell) < 0].$$

This simulator is unbiased and smooth in the model parameters. Note that construction of the GHK simulator requires only draws from truncated univariate normals and evaluation of univariate integrals. To draw from a truncated univariate normal is quite simple, since, if a standard normal random variate η is desired such that $a < \eta < b$, one need only form $\eta = F^{-1}[(F(b) - F(a))U + F(a)]$ where $F(\cdot)$ is the standard normal distribution function and U is a uniform random variate on [0,1].

4.2 Simulated Maximum Likelihood (SML)

Albright, Lerman, and Manski (1977) and Lerman and Manski (1981) proposed maximum likelihood using (2.3) with $\hat{P}_F(\cdot)$ in lieu of $P(\cdot)$,

$$\hat{L}_N(\beta^*,\gamma^*,\Sigma^*) = N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{J} d_{ij}\ln[\hat{P}_F(j\,|\,\beta^*,\gamma^*,\Sigma^*,X_i^*,Z_i^*)].$$

If $M/N^{1/2} \to \infty$ as $N \to \infty$, then this estimator is consistent (see Lee 1992 or McFadden and Ruud 1992). The same $M/N^{1/2} \to \infty$ condition guarantees consistency of the maximum likelihood estimator based on $\hat{P}_{GHK}$ in lieu of $P(\cdot)$, and of the maximum likelihood estimator based on $\hat{P}_{KS}(\cdot)$ in lieu of $P(\cdot)$ provided that $\rho \to 0$ appropriately as $N \to \infty$. None of the estimators is consistent in N (with M fixed) since they all provide biased evaluations of $\ln P(\cdot)$.

SML based on the frequency simulator has two undesirable characteristics. First, $\hat{L}_N$ will be discontinuous in β^*, γ^*, and Σ^*, precluding the use of the gradient methods for optimization and statistical inference. Second, a simulated choice probability of zero precludes construction of $\hat{L}_N$. If the frequency simulator is replaced by a smooth probability simulator that is bounded away from zero, such as the kernel-smoothed or GHK simulator, these problems are avoided. However, $\hat{L}_N$ still provides a biased evaluation of L_N, for fixed M. An important open question, which we examine in this paper, is whether or not this imparts a substantial bias to the simulated maximum likelihood estimator.

4.3 Method of Simulated Moments (MSM)

McFadden (1989) and Pakes and Pollard (1989) observe that the solution of the simulated moment conditions

$$N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{J} W_{ij}[d_{ij} - \hat{P}(j\,|\,\beta^*,\gamma^*,\Sigma^*,X_i,Z_i)] = 0,$$

where W_{ij} is a set of instruments and $\hat{P}(\cdot)$ denotes any unbiased probability simulator, will produce estimates of β^*, γ^*, and Σ^* that are consistent in N, with fixed M. Use of biased probability simulators also results in consistent estimates if the bias is $o(N^{-1/2})$. In the case of $\hat{P}_{KS}$, this condition requires that $\rho \rightarrow 0$ appropriately as $N \rightarrow \infty$. Since simulation is used to obtain the moment conditions, McFadden (1989) called this estimator the Method of Simulated Moments.

McFadden (1989) points out that the instruments $(\partial\hat{P}(\cdot)/\partial\theta)/\hat{P}(\cdot)$ are asymptotically optimal if $M \rightarrow \infty$, where θ is the vector of free parameters in β^*, γ^*, and Σ^*. Consistency also requires that the draws used to form $\partial\hat{P}(\cdot)/\partial\theta$ be independent of those used to form $d_{ij} - \hat{P}(j\,|\,\beta^*,\gamma^*,\Sigma^*, X_i,Z_i)$.

5. Bayesian Inference Using the Gibbs Sampler

Bayesian inference using the Gibbs sampler (Gelfand and Smith 1990) and data augmentation (Tanner and Wong 1987) has been applied to the multinomial probit model by at least two other sets of investigators (Albert and Chib 1992, McCulloch and Rossi 1992). Here, we describe the essentials of the method.

In conjunction with the log-likelihood function (2.3) posterior densities corresponding to various priors are easy to construct as formal expressions. The essential difficulty in applying any of these expressions is integrating over the unobserved components u_i^*, and in Bayesian inference there is the further complication of integrating over the parameters β^*, γ^*, and Σ^*. The Gibbs sampler with data augmentation resolves both difficulties in a systematic way. We describe the method in turn for three variance structures discussed in Section 2.

The third variance structure is simplest: Σ is scalar, and hence $\Sigma^* = 0.5I + 0.5ee'$, where e is a $(J-1) \times 1$ vector of units. Let the priors on β^* and γ^* be diffuse: $\pi_{\beta^*}(\beta^*) \propto$ constant, $\pi_{\gamma^*}(\gamma^*) \propto$ constant. The essence of Gibbs sampling and data augmentation is that, under weak conditions widely satisfied by econometric models (including this one), successive sampling from conditional distributions produces a Markov chain which converges in distribution to the posterior distribution (Tierney 1991, McCulloch and Rossi 1992, and Geweke 1992). So, to implement the Gibbs sampler, start by choosing initial values for the model parameters β^* and γ^*. Then, in each iteration of the algorithm there are two steps:

1. Conditional on β^*, γ^*, and Σ^*, the distribution of u_i^* is truncated $(J-1)$-variate normal. If the observed choice is j, the truncations are created by the linear restrictions $u_{ik}^* - u_{ij}^* \leq 0$ for $k = 1, \ldots, J$. A simple method for drawing from a trun-

cated multivariate normal distribution described in Geweke (1991) can be applied. Essentially, this method exploits the fact that conditional on β^*, γ^*, and the u_{ik}^* for $k \neq l$, u_{il}^* has a truncated univariate normal distribution, from which the construction of synthetic random variables is trivial.

2. Conditional on u_i^*, (2.2) is a seemingly unrelated regressions model (Zellner 1962), and the posterior distribution of β^* and γ^* in this model is joint normal with mean and variance given by familiar generalized least squares expressions (Zellner 1971). While these expressions may be used directly, they require the inversion of a symmetric, positive-definite $(m+r) \times (m+r)$ matrix. In the experiments undertaken for this study the matrix $[X_i^* | Z_i^*]$ of covariates is sparse, with many more zero than nonzero entries. Techniques described in Geweke, Keane, and Runkle (1994) for such systems were employed for the computations reported in this paper, greatly increasing speed and reducing storage requirements.

Beginning from the arbitrary initial values for β^* and γ^*, the Gibbs sampling-data augmentation algorithm alternates between Steps 1 and 2. At the ℓth iteration drawings $\beta^{*(\ell)}$ and $\gamma^{*(\ell)}$ are produced. As the number of iterations grows large the sequence $\{\beta^{*(\ell)},\gamma^{*(\ell)}\}$ converges in distribution to the posterior distribution. Therefore, the sequence $\{g(\beta^{*(\ell)},\gamma^{*(\ell)})\}$ converges in distribution to the posterior distribution of the function $g(\beta^*,\gamma^*)$. The assessment of convergence, and of the numerical accuracy of approximations to posterior moments, is an important task to which contributions are currently being made (see Geweke 1992, Schervish and Carlin 1992, and Zellner and Min 1992).

In the first and least restrictive specification Σ^* must be renormalized in some way. We accomplish this through the diffuse but proper prior specification,

$$\Sigma^* \sim \mathrm{IW}(0.5(\mathrm{I}_{J-1}+ee'), 1)$$

while retaining flat, improper priors for β^* and γ^* (This approach is similar to McCulloch and Rossi (1994), except that they also use proper priors for the coefficients.) This modification leads to a third step in each iteration:

3. Conditional on β^*, γ^*, and the u_i^*, the distribution of Σ^* is inverted Wishart. An appropriate drawing may be made by generating the synthetic random variables from the inverted Wishart distribution as described in Geweke (1988) and then normalizing Σ^* and the coefficients β^* and γ^*.

Of course, the synthetic random variables in Steps 1 and 2 are then drawn conditional on Σ^* from Step 3 of the previous iteration, in lieu of the fixed $\Sigma^* = 0.5\mathrm{I} + 0.5ee'$ used for the first variance structure. For comparison with the classical results, we record as functions of interest in each iteration the normalized values $\beta^{*(\ell)}(\sigma_{11}^{*(\ell)})^{-1/2}$ and $\gamma^{*(\ell)}(\sigma_{11}^{*(\ell)})^{-1/2}$ in lieu of $\beta^{*(\ell)}$ and $\gamma^{*(\ell)}$, and the Cholesky decomposition of $\Sigma^*(\sigma_{11}^{*(\ell)})^{-1/2}$ rather than Σ^*.

The other variance structure taken up in the experiments specifies that the variance matrix Σ in the unnormalized model (2.1) is diagonal. This induces the variance structure (3.6), which is equivalent to the factor model:

$$\epsilon_{ij}^* = \alpha f_i + \zeta_{ij},$$

where $\alpha = 0.5^{1/2}$, the latent variables $(f_i, \zeta_{i1}, \ldots, \zeta_{i,J-1})$ are mutually independent with $f_i \sim \mathrm{IIDN}(0,1)$, $\zeta_i \sim \mathrm{IIDN}(0,\Delta)$, and $\Delta = \mathrm{Diag}(\delta_1, \ldots, \delta_{J-1})$. Since this model is identified we may adopt the conventional diffuse priors $\pi_{\delta_i}(\delta_i) \propto \delta_i^{-1}$, $\pi_{\beta^*}(\beta^*) \propto$ constant, and $\pi_{\gamma^*}(\gamma^*) \propto$ constant. Thus (2.2) becomes:

$u_i^* - \alpha f_i = X_i^* \beta^* + Z_i^* \gamma^* + \zeta_i$, (5.1).

The functions of interest are $\beta^{*(\ell)}(\sigma_{11}^{*(\ell)})^{-1/2}$, $\gamma^{*(\ell)}(\sigma_{11}^{*(\ell)})^{-1/2}$ and, for comparison with (3.6), $\xi_j = \delta_j - 0.5$. The foregoing procedures may still be applied, with modification as follows.

1. Drawings of u_i^* conditional on β^*, γ^*, $f = (f_1,\dots,f_n)'$, and the variance matrix Δ are made as before.

1.5. Following Step 1 and before Step 2, drawings from the distribution of the f_i conditional on u_i^*, β^*, γ^*, α, and Δ are made from the appropriate distribution,

$$f_i \sim N[e\alpha(\alpha^2 ee' + \Delta)^{-1}(u_i^* - X_i^* \beta^* - Z_i^* \gamma^*),\ I - \alpha^2 e(\alpha^2 ee' + \Delta)^{-1} e'].$$

2. Conditional on the f_i, u_i^*, and Δ, (5.1) is a seemingly unrelated regressions model. The conditional distribution of β^* and γ^* is therefore joint normal with mean and variance given by the generalized least squares formulas.

3. Conditional on f_i and u_i^*, the posterior distribution of the δ_j are independent inverted gamma (Zellner 1971), from which synthetic random variables may be constructed in trivial fashion as described in Geweke (1986).

All results are reported for $m = 10{,}000$ iterations of the Gibbs sampling-data augmentation algorithm. In every case, iterations began with the starting values $\beta^* = 0$, $\gamma^* = 0$, $\Sigma^* = 0.5ee' + 0.5I$. Experimentation with starting values much further from population values showed that draws from the posterior distribution converged to the neighborhood of the population values within 100 iterations. Accordingly, only the first 200 iterations were discarded to reduce sensitivity to initial conditions. The posterior mean of a function of interest $\bar{g} \equiv E[g(\beta^*, \gamma^*, \Sigma^*)]$ is approximated by the corresponding moment from the sample generated by the algorithm, $\bar{g}_m \equiv m^{-1} \sum_{\ell=1}^{m} g(\beta^{*(\ell)},\gamma^{*(\ell)},\Sigma^{*(\ell)})$. The posterior variance is approximated by $m^{-1} \sum_{\ell=1}^{m} [g(\beta^{*(\ell)},\gamma^{*(\ell)},\Sigma^{*(\ell)}) - \bar{g}_m]^2$, and the posterior standard deviation is approximated by the square root of this expression.

6. Results of the Experiments

A. Experiment One

In the first experiment, each of the nine models described in Section 2 was estimated using 10 artificial data sets, whose construction is also described in Section 2. The purpose of this experiment is to compare four estimators of these nine models:

1. Posterior means using the Gibbs sampling-data augmentation algorithm with $m = 10{,}000$ iterations;
2. Method of simulated moments using the GHK probability simulator with $M = 30$ draws to simulate the choice probabilities and the derivatives needed to form the optimal weights, and using Gauss-Newton iterations to solve the simulated moment conditions;
3. Method of simulated moments using the KS probability simulator with $M = 100$ draws to simulate the choice probabilities and the optimal weights, with both $\rho = 0.10$ and $\rho = 0.20$, again using Gauss-Newton iterations for solution;
4. Simulated maximum likelihood using the GHK probability simulator with $M = 30$ draws to simulate the choice probabilities, and using BHHH iterations to maximize the simulated log-likelihood function.

We did not employ the frequency simulator for reasons discussed in Section 4.1. In the case of MSM based on the GHK simulator, the value of M was chosen such that increases in M had negligible effect on the point estimates. In the cases of MSM based on the KS simulator and SML based on the GHK simulator, M was chosen to give computation times close to that for MSM based on GHK. In the case of Gibbs sampling the choice m = 10,000 has become fairly standard and the resultant accuracy as indicated by the numerical standard errors (computed as described in Geweke (1992) but not reported here) appears to be quite acceptable.

In tables 1–9 we report the means over the 10 artificial data sets of the point estimates or posterior means, and also of the asymptotic standard errors (ASE) or posterior standard deviations (PSD), for all parameters in each model. Note that for the unrestricted Σ^* specification these parameters include A, and in the diagonal variance models they include the ξ_i.

Several interesting patterns arise when one compares the performance of different estimators and when one compares the general performance of the estimators across model specifications and across different groups of parameters. We will now compare the performance of each method for each set of parameters in each model, focussing primarily on how the MSEs for those sets of parameters compare across methods and models.

Table 1 contains results for the unrestricted model, consisting of coefficient specification (3.1) and variance specification (3.2). For the γ^*_{j1}, the MSE for MSM-KS with $\rho = 0.10$ or 0.20 and for SML-GHK are all roughly 2.5 to 3 times greater than those for MSM-GHK. The MSE for Bayesian inference are roughly 10%–20% smaller than those for MSM-GHK, and are best in 7 of 7 cases. For MSM-GHK there is close agreement between the MSE and the mean ASE, suggesting that the asymptotic distribution theory provides a good approximation to the

Table 1: Coefficient specification (3.1), Unrestricted Σ^* specification

		Bayesian Inference			*MSM-GHK*			*SML-GHK*			*MSM-KS*					
											ρ = .10			ρ = .20		
θ	DGP	$\bar{\bar{\theta}}$	MSE	$\overline{\text{PSD}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$
β^*_{11}	0.000	0.041	0.073	0.106	0.032	0.143	0.124	0.042	0.115	0.084	0.021	0.168	0.199	0.013	0.107	0.151
β^*_{21}	0.000	-0.001	0.077	0.114	-0.007	0.102	0.119	-0.030	0.084	0.097	0.000	0.235	0.191	-0.032	0.158	0.149
β^*_{31}	0.000	0.000	0.082	0.116	0.029	0.121	0.117	-0.048	0.119	0.101	-0.071	0.274	0.173	-0.081	0.165	0.140
β^*_{41}	0.000	-0.001	0.088	0.110	-0.007	0.141	0.118	-0.051	0.099	0.101	0.029	0.167	0.172	-0.016	0.104	0.140
β^*_{51}	0.000	-0.013	0.070	0.119	-0.006	0.116	0.116	-0.088	0.124	0.107	-0.017	0.183	0.165	-0.059	0.134	0.136
β^*_{61}	0.000	0.017	0.147	0.112	0.025	0.191	0.117	-0.044	0.144	0.101	0.014	0.165	0.166	0.013	0.119	0.136
β^*_{12}	0.000	0.009	0.029	0.029	0.009	0.030	0.029	0.003	0.023	0.027	0.008	0.032	0.030	0.007	0.029	0.028
β^*_{22}	0.000	-0.009	0.037	0.031	-0.009	0.040	0.029	-0.014	0.035	0.028	-0.010	0.029	0.029	-0.012	0.032	0.027
β^*_{32}	0.000	-0.013	0.034	0.030	-0.011	0.031	0.029	-0.012	0.035	0.028	-0.003	0.027	0.028	-0.008	0.030	0.027
β^*_{42}	0.000	0.005	0.042	0.030	0.005	0.041	0.029	0.003	0.043	0.028	0.010	0.040	0.029	0.006	0.042	0.028
β^*_{52}	0.000	0.006	0.026	0.030	0.006	0.026	0.029	0.006	0.026	0.029	0.004	0.036	0.029	0.006	0.032	0.027
β^*_{62}	0.000	0.000	0.028	0.030	0.001	0.028	0.029	0.006	0.028	0.029	0.012	0.032	0.029	0.009	0.029	0.027
γ^*_{11}	0.707	0.701	0.052	0.055	0.711	0.074	0.060	0.657	0.232	0.045	0.633	0.242	0.086	0.693	0.242	0.065
γ^*_{21}	0.707	0.706	0.046	0.077	0.718	0.058	0.083	0.682	0.234	0.069	0.629	0.264	0.121	0.701	0.255	0.094
γ^*_{31}	0.707	0.714	0.069	0.080	0.716	0.083	0.082	0.693	0.241	0.071	0.668	0.270	0.112	0.728	0.264	0.089
γ^*_{41}	0.707	0.718	0.072	0.077	0.729	0.102	0.082	0.705	0.243	0.071	0.624	0.252	0.114	0.703	0.254	0.091
γ^*_{51}	0.707	0.724	0.064	0.082	0.728	0.079	0.081	0.726	0.250	0.075	0.648	0.255	0.110	0.726	0.264	0.088
γ^*_{61}	0.707	0.704	0.085	0.077	0.706	0.111	0.082	0.699	0.245	0.071	0.635	0.264	0.109	0.687	0.252	0.088
γ^*_{71}	-0.707	-0.719	0.056	0.054	-0.726	0.078	0.060	-0.678	0.234	0.045	-0.632	0.236	0.083	-0.692	0.241	0.063
a^*_{21}	0.500	0.517	0.100	0.112	0.507	0.134	0.128	0.449	0.189	0.082	0.435	0.280	0.186	0.401	0.254	0.152
a^*_{22}	0.866	0.906	0.084	0.106	0.829	0.266	0.111	0.846	0.292	0.089	0.776	0.308	0.157	0.836	0.314	0.137
a^*_{31}	0.500	0.515	0.101	0.109	0.527	0.130	0.123	0.488	0.207	0.083	0.443	0.297	0.187	0.444	0.243	0.151
a^*_{32}	0.289	0.355	0.099	0.108	0.342	0.110	0.120	0.279	0.120	0.079	0.196	0.284	0.176	0.285	0.197	0.140
a^*_{33}	0.816	0.818	0.094	0.100	0.821	0.122	0.104	0.796	0.284	0.088	0.684	0.344	0.141	0.748	0.309	0.127
a^*_{41}	0.500	0.513	0.094	0.106	0.495	0.125	0.121	0.441	0.175	0.085	0.429	0.258	0.188	0.427	0.257	0.150
a^*_{42}	0.289	0.277	0.104	0.107	0.312	0.153	0.121	0.267	0.113	0.078	0.330	0.207	0.181	0.318	0.210	0.145
a^*_{43}	0.204	0.187	0.108	0.105	0.149	0.114	0.114	0.149	0.105	0.078	0.085	0.255	0.171	0.079	0.203	0.137
a^*_{44}	0.791	0.793	0.086	0.093	0.805	0.116	0.102	0.796	0.274	0.087	0.630	0.281	0.140	0.689	0.265	0.127
a^*_{51}	0.500	0.563	0.108	0.113	0.556	0.126	0.120	0.507	0.189	0.087	0.497	0.266	0.200	0.525	0.245	0.152
a^*_{52}	0.289	0.286	0.089	0.110	0.264	0.087	0.119	0.240	0.121	0.083	0.248	0.144	0.191	0.273	0.149	0.145
a^*_{53}	0.204	0.201	0.103	0.111	0.225	0.099	0.112	0.189	0.102	0.081	0.169	0.170	0.168	0.170	0.125	0.133
a^*_{54}	0.158	0.139	0.107	0.108	0.128	0.098	0.108	0.146	0.081	0.077	0.187	0.163	0.156	0.137	0.153	0.127
a^*_{55}	0.775	0.764	0.082	0.097	0.782	0.087	0.098	0.802	0.284	0.091	0.627	0.317	0.136	0.702	0.283	0.123
a^*_{61}	0.500	0.503	0.102	0.107	0.503	0.184	0.121	0.485	0.197	0.087	0.530	0.262	0.191	0.488	0.226	0.152
a^*_{62}	0.289	0.294	0.109	0.109	0.305	0.147	0.120	0.267	0.120	0.081	0.300	0.238	0.171	0.275	0.195	0.142
a^*_{63}	0.204	0.231	0.129	0.103	0.242	0.151	0.112	0.232	0.130	0.078	0.146	0.176	0.166	0.150	0.177	0.134
a^*_{64}	0.158	0.152	0.080	0.106	0.132	0.103	0.108	0.158	0.108	0.076	0.102	0.167	0.167	0.119	0.148	0.129
a^*_{65}	0.129	0.160	0.131	0.100	0.143	0.155	0.103	0.178	0.124	0.074	0.112	0.186	0.148	0.103	0.148	0.124
a^*_{66}	0.764	0.711	0.135	0.090	0.719	0.157	0.098	0.741	0.265	0.084	0.584	0.296	0.136	0.623	0.275	0.123

Note: $\theta \equiv$ parameter, DGP $\equiv$ data generating value, $\bar{\bar{\theta}} \equiv$ average parameter estimate,
MSE $\equiv$ root mean square error, $\overline{\text{PSD}} \equiv$ average posterior standard deviation,
$\overline{\text{ASE}} \equiv$ average asymptotic standard error.

small sample distribution of the MSM estimates. For Bayesian inference there is also close agreement between the MSE and the mean PSD. However, for MSM-KS and SML-GHK the mean ASE greatly underestimate the MSE, suggesting that these methods produce standard errors for the γ^*_{j1} estimates that are biased downward by factors of 2.5 to 3.

For the β^*_{j2} the MSE are similar across all 5 methods. MSM-KS ($\rho = 0.10$) is best in 3 of 6 cases, SML-GHK is best in 1 of 6, and MSM-GHK, SML-GHK and Bayesian inference are tied for best in 2 of 6. For the β^*_{j1} the MSE for MSM-KS ($\rho = 0.10$) are generally the largest. There is considerable improvement in going to MSM-KS ($\rho = 0.20$), which produces MSE roughly comparable to those for MSM-GHK. The MSE for Bayesian Inference are lowest in 5 of 6 cases, and the MSE for SML-GHK are generally second best.

For the a_{ij} the MSE for MSM-KS ($\rho = 0.10$) are generally about 2 times greater than those for MSM-GHK. Those for MSM-KS ($\rho = 0.20$) are somewhat better but are still generally much larger than for MSM-GHK. For SML-GHK, the MSE for the diagonal elements a_{ii} are generally about 2 times larger than those for MSM-GHK. The MSE for Bayesian Inference are generally about 20% smaller than those for MSM-GHK, and are smallest in 15 of 20 cases.

We see in table 1 that the ranking of methods in terms of MSE depends on the type of parameter considered, but the most striking feature of the results is the poor performance of MSM-KS and SML-GHK for the γ^*_{j1} parameters. Overall, the performance of MSM-GHK and Bayesian inference appear to dominate, with Bayesian inference getting the slight edge (especially in terms of MSE for the γ^*_{j1} and β^*_{j1} parameters).

Despite the appearance of certain problems for specific methods for specific parameters, we find the overall precision of the estimates and posterior means in table 1 somewhat surprising, given that the unrestricted model contains 20 covariance matrix parameters and that such parameters are notoriously difficult to estimate in discrete choice models (see Keane 1992). This difficulty arises due to the loss of information involved in only observing discrete outcomes rather than the underlying continuous latent variables that determine outcomes.

In table 2 we impose the restriction that the γ^*_{j1} are equal for all j. This is a restriction that one would often impose in practice. For example, the z_{ij} may be prices, and the restriction may correspond to imposing homogeneity of degree zero on demand. The ranking of methods in terms of MSE for γ^*_{11} is (1) Bayesian inference at 0.043, (2) SML-GHK at 0.057, (3) MSM-GHK at 0.064, (4) MSM-KS ($\rho = 0.10$) at 0.075, and (5) MSM-KS ($\rho = 0.20$) at 0.091. The improvement of these MSEs from those in table 1 is dramatic. The MSEs for γ^*_{11} are often 2 to 3 times smaller than those for the individual γ^*_{j1} in table 1.

For the β^*_{j2} the drop in MSEs is slight, and the MSE are rather similar across methods. MSM-GHK is best or tied for best in 3 of 6 cases, followed by SML-GHK and MSM-KS ($\rho = 0.10$) at 2 of 6 and Bayesian Inference at 1 of 6.

For the β^*_{j1} the drop in MSEs that results from imposing equality of the γ^*_{j1} is dramatic. For Bayesian inference the drops range from 10% to 75%. For SML-GHK they fall by roughly a factor of 2, and for MSM-GHK and MSM-KS they fall by factors of 2 to 3. The MSE for MSM-KS ($\rho = 0.10$) are worst in 6 of 6 cases. Those for MSM-KS ($\rho = 0.20$) are 10% to 20% smaller, but are in all cases 2nd worst. The MSE for Bayesian inference are best or tied for best in 5 of 6 cases, but are generally less than 10% smaller than those for MSM-GHK and SML-GHK.

For the a_{ij} the drops in MSEs are also often dramatic. For MSM-KS the MSE generally fall by factors of 1.5 to 3, the improvement for Bayesian inference is generally about 10%, and that for MSM-GHK is generally about 10% to 30%. For SML-GHK, the MSE for the diagonal elements a_{ii} improve by roughly a factor of 3, and those for the other elements often improve by factors of 2. The MSEs for MSM-KS ($\rho = 0.10$) are worst in 11 of 20 cases while those for MSM-KS ($\rho = 0.20$) are worst in 9 of 20 cases. The MSEs for the MSM-KS methods are often about 50% to 100% greater than those for MSM-GHK. The MSE for SML-GHK are best in 11 of 20 cases, Bayesian inference is best in 8 of 20 cases, and MSM-GHK is best in 1 of 20 cases.

Overall, the method that improves most with the γ_{j1}^* equality restriction is SML-GHK, due to the large drop in the MSE for γ_{11}^* relative to those for the γ_{j1}^* in table 1, and the large drops in the MSEs for the a_{ij}. In the second table, Bayesian inference again has a slight overall edge in performance in terms of MSEs, with MSM-GHK and SML-GHK difficult to choose between: the former tends to have slightly better MSEs for the β^* while the later does better for the a_{ij}. Again, the MSM-KS point estimates clearly have the largest MSEs.

In table 3 we impose the further restriction that the β_{jk}^* are zero for all j and k = 1, 2. This restriction is perhaps harder to justify than that in table 2 in terms of basic theory. However, it is common in marketing applications to construct models where only brand attributes determine choices (perhaps with the coefficients on attributes depending on household characteristics), and where the error structure arises from households heterogeneous preferences for unobserved brand attributes (see, e.g., Elrod and Keane 1994). Thus, a model without β_{jk}^* is in fact of independent interest.

With this restriction, the MSE for the γ_{11}^* fall by about 20% from those in table 2. The ranking of methods in terms of MSE for γ_{11}^* is again (1) Bayesian inference at 0.034, (2) SML-

Table 2: Coefficient specification (3.4), Unrestricted Σ^* specification

		Bayesian Inference			*MSM-GHK*			*SML-GHK*			*MSM-KS*					
											ρ = .10			ρ = .20		
θ	DGP	$\bar{\theta}$	MSE	$\overline{\text{PSD}}$	$\bar{\theta}$	MSE	$\overline{\text{ASE}}$	$\bar{\theta}$	MSE	$\overline{\text{ASE}}$	$\bar{\theta}$	MSE	$\overline{\text{ASE}}$	$\bar{\theta}$	MSE	$\overline{\text{ASE}}$
β^*_{11}	0.000	0.012	0.051	0.050	0.014	0.054	0.048	0.010	0.051	0.048	0.024	0.080	0.062	0.019	0.069	0.056
β^*_{21}	0.000	-0.019	0.037	0.051	-0.019	0.043	0.051	-0.024	0.040	0.051	-0.003	0.066	0.065	-0.017	0.055	0.059
β^*_{31}	0.000	-0.007	0.072	0.052	-0.006	0.068	0.050	-0.020	0.074	0.051	-0.011	0.099	0.063	-0.022	0.086	0.058
β^*_{41}	0.000	-0.005	0.051	0.051	-0.003	0.053	0.050	-0.021	0.059	0.051	-0.001	0.077	0.062	-0.008	0.071	0.057
β^*_{51}	0.000	-0.006	0.064	0.051	-0.003	0.066	0.050	-0.024	0.070	0.051	0.002	0.092	0.061	-0.010	0.082	0.057
β^*_{61}	0.000	-0.003	0.048	0.051	0.000	0.048	0.050	-0.023	0.052	0.051	0.003	0.065	0.062	-0.001	0.056	0.057
β^*_{12}	0.000	0.008	0.031	0.029	0.009	0.029	0.029	0.008	0.028	0.029	0.013	0.036	0.033	0.013	0.034	0.031
β^*_{22}	0.000	-0.009	0.039	0.029	-0.007	0.040	0.029	-0.008	0.037	0.030	-0.007	0.035	0.032	-0.007	0.038	0.030
β^*_{32}	0.000	-0.011	0.033	0.029	-0.010	0.031	0.029	-0.011	0.033	0.030	-0.005	0.027	0.032	-0.008	0.030	0.030
β^*_{42}	0.000	0.005	0.042	0.029	0.005	0.039	0.029	0.005	0.041	0.030	0.014	0.041	0.032	0.009	0.043	0.031
β^*_{52}	0.000	0.006	0.026	0.029	0.006	0.025	0.029	0.008	0.025	0.031	0.004	0.035	0.032	0.008	0.030	0.030
β^*_{62}	0.000	0.002	0.026	0.029	0.002	0.026	0.029	0.002	0.028	0.031	0.010	0.031	0.032	0.006	0.029	0.030
γ^*_{11}	0.707	0.710	0.043	0.042	0.719	0.064	0.047	0.738	0.057	0.039	0.725	0.075	0.068	0.774	0.091	0.051
a^*_{21}	0.500	0.502	0.127	0.088	0.490	0.132	0.105	0.497	0.093	0.074	0.485	0.144	0.157	0.460	0.202	0.129
a^*_{22}	0.866	0.911	0.087	0.084	0.915	0.123	0.092	0.931	0.103	0.076	0.928	0.139	0.137	0.939	0.140	0.117
a^*_{31}	0.500	0.490	0.090	0.093	0.511	0.084	0.100	0.516	0.093	0.074	0.505	0.139	0.155	0.502	0.164	0.127
a^*_{32}	0.289	0.353	0.091	0.096	0.340	0.103	0.117	0.326	0.096	0.084	0.248	0.166	0.189	0.331	0.161	0.151
a^*_{33}	0.816	0.817	0.090	0.071	0.823	0.113	0.078	0.858	0.103	0.066	0.802	0.125	0.116	0.800	0.134	0.099
a^*_{41}	0.500	0.493	0.081	0.088	0.477	0.094	0.099	0.458	0.088	0.075	0.460	0.201	0.158	0.442	0.188	0.126
a^*_{42}	0.289	0.284	0.088	0.104	0.305	0.151	0.118	0.310	0.081	0.082	0.365	0.197	0.193	0.374	0.205	0.155
a^*_{43}	0.204	0.191	0.105	0.101	0.151	0.118	0.112	0.169	0.084	0.081	0.128	0.192	0.183	0.101	0.176	0.147
a^*_{44}	0.791	0.786	0.062	0.068	0.791	0.078	0.071	0.841	0.083	0.063	0.761	0.085	0.104	0.770	0.073	0.091
a^*_{51}	0.500	0.538	0.086	0.086	0.536	0.076	0.097	0.530	0.066	0.076	0.538	0.155	0.167	0.583	0.162	0.126
a^*_{52}	0.289	0.288	0.081	0.104	0.265	0.091	0.116	0.276	0.063	0.085	0.310	0.108	0.201	0.307	0.129	0.155
a^*_{53}	0.204	0.208	0.117	0.103	0.226	0.105	0.110	0.228	0.085	0.081	0.216	0.131	0.181	0.209	0.100	0.143
a^*_{54}	0.158	0.141	0.091	0.102	0.134	0.078	0.107	0.159	0.055	0.079	0.162	0.128	0.169	0.149	0.137	0.138
a^*_{55}	0.775	0.752	0.067	0.064	0.770	0.078	0.065	0.819	0.089	0.060	0.735	0.131	0.094	0.743	0.108	0.084
a^*_{61}	0.500	0.487	0.099	0.089	0.490	0.111	0.098	0.505	0.091	0.078	0.554	0.158	0.158	0.534	0.132	0.128
a^*_{62}	0.289	0.295	0.104	0.105	0.307	0.150	0.117	0.320	0.090	0.086	0.358	0.232	0.185	0.348	0.198	0.154
a^*_{63}	0.204	0.236	0.128	0.101	0.249	0.159	0.110	0.267	0.118	0.082	0.203	0.137	0.180	0.173	0.179	0.145
a^*_{64}	0.158	0.142	0.083	0.104	0.142	0.091	0.106	0.176	0.085	0.080	0.092	0.155	0.180	0.133	0.134	0.140
a^*_{65}	0.129	0.157	0.129	0.098	0.148	0.151	0.102	0.184	0.107	0.077	0.118	0.162	0.166	0.096	0.158	0.137
a^*_{66}	0.764	0.719	0.082	0.063	0.730	0.095	0.061	0.797	0.083	0.058	0.713	0.106	0.089	0.721	0.099	0.079

Note: $\theta \equiv$ parameter, DGP $\equiv$ data generating value, $\bar{\theta} \equiv$ average parameter estimate,
MSE $\equiv$ root mean square error, $\overline{\text{PSD}} \equiv$ average posterior standard deviation,
$\overline{\text{ASE}} \equiv$ average asymptotic standard error.

GHK at 0.043, (3) MSM-GHK at 0.051, (4) MSM-KS (ρ = 0.10) at 0.064, and (5) MSM-KS (ρ = 0.20) at 0.076.

For the a_{ij} the MSEs for SML-GHK fall by about 0% to 20%, while those for Bayesian Inference, MSM-GHK and MSM-KS typically fall by about 0% to 10%. SML-GHK is best in 13 of 20 cases, Bayesian inference is best in 6 of 20, and MSM-GHK is best in 1 of 20. MSM-KS (ρ = 0.10) is worst or tied for worst in 12 of 20, while MSM-KS (ρ = 0.20) is worst or tied for worst in 9 of 20.

Overall, in table 3, it is difficult to choose a best method. Bayesian inference performs best for γ^*_{11}, while SML-GHK has a slight edge for the a_{ij}. MSM-GHK is clearly third for both but is not far behind. SML-GHK is the method most helped by the restriction, since in the second experiment it produced MSEs for the β^*_{jk} that were larger than those obtained by MSM-GHK and Bayesian inference. Again, MSM-KS is clearly dominated by other methods.

In table 4 we return to the general model and impose the restriction that the covariance matrix of the untransformed model is diagonal. This corresponds to a model where the only unobservables are unique attributes of alternatives for which households have heterogeneous preferences, and where the alternatives have different levels of the unique attributes.

Comparing the results in table 4 to those in table 1, we see that imposing diagonality on Σ leads in some cases to dramatic reductions in the MSE for the γ^*_{j1}. These fall by factors of about 1.5 to 2 for MSM-GHK, 2.5 to 5 for SML-GHK, 2 to 3 for MSM-KS (ρ = 0.20), and 2 to 5 for MSM-KS (ρ = 0.10). For Bayesian inference, however, the MSE rise more often than not. The MSE for MSM-GHK are best in 7 of 7 cases, and it is clear that MSM-GHK dominates along this dimension. The MSE for Bayesian inference and SML-GHK are similar. Those for MSM-KS are generally about 50% to 300% greater than those for MSM-GHK, but

Table 3: Coefficient specification (3.5), Unrestricted Σ^* specification

		Bayesian Inference			*MSM-GHK*			*SML-GHK*			*MSM-KS*					
											$\rho = .10$			$\rho = .20$		
θ	DGP	$\bar{\theta}$	MSE	$\overline{\text{PSD}}$	$\bar{\theta}$	MSE	$\overline{\text{ASE}}$	$\bar{\theta}$	MSE	$\overline{\text{ASE}}$	$\bar{\theta}$	MSE	$\overline{\text{ASE}}$	$\bar{\theta}$	MSE	$\overline{\text{ASE}}$
γ^*_{11}	0.707	0.703	0.034	0.037	0.709	0.051	0.042	0.723	0.043	0.034	0.717	0.064	0.063	0.762	0.076	0.046
a^*_{21}	0.500	0.500	0.118	0.080	0.483	0.121	0.098	0.490	0.091	0.064	0.474	0.129	0.150	0.447	0.197	0.121
a^*_{22}	0.866	0.892	0.074	0.078	0.900	0.118	0.086	0.915	0.094	0.068	0.917	0.139	0.134	0.919	0.130	0.111
a^*_{31}	0.500	0.481	0.076	0.079	0.506	0.070	0.094	0.509	0.077	0.065	0.494	0.120	0.148	0.490	0.137	0.119
a^*_{32}	0.289	0.352	0.088	0.098	0.348	0.099	0.117	0.340	0.098	0.081	0.255	0.157	0.188	0.338	0.151	0.150
a^*_{33}	0.816	0.805	0.071	0.060	0.814	0.099	0.069	0.839	0.080	0.055	0.783	0.102	0.109	0.778	0.113	0.090
a^*_{41}	0.500	0.482	0.078	0.076	0.471	0.089	0.093	0.452	0.076	0.066	0.449	0.190	0.149	0.433	0.185	0.118
a^*_{42}	0.289	0.294	0.094	0.097	0.311	0.148	0.117	0.321	0.082	0.079	0.371	0.190	0.191	0.376	0.199	0.154
a^*_{43}	0.204	0.184	0.093	0.095	0.150	0.117	0.112	0.172	0.086	0.078	0.124	0.192	0.182	0.106	0.170	0.146
a^*_{44}	0.791	0.780	0.052	0.056	0.783	0.066	0.060	0.822	0.066	0.050	0.751	0.090	0.094	0.754	0.093	0.078
a^*_{51}	0.500	0.526	0.075	0.076	0.531	0.073	0.092	0.523	0.066	0.067	0.534	0.138	0.158	0.571	0.144	0.119
a^*_{52}	0.289	0.290	0.084	0.102	0.273	0.093	0.116	0.289	0.067	0.081	0.306	0.107	0.200	0.315	0.128	0.154
a^*_{53}	0.204	0.205	0.106	0.097	0.222	0.102	0.110	0.227	0.082	0.078	0.220	0.130	0.180	0.212	0.108	0.142
a^*_{54}	0.158	0.147	0.095	0.097	0.133	0.077	0.107	0.163	0.048	0.075	0.167	0.134	0.168	0.145	0.134	0.137
a^*_{55}	0.775	0.743	0.047	0.052	0.763	0.058	0.053	0.796	0.053	0.046	0.721	0.126	0.082	0.724	0.094	0.069
a^*_{61}	0.500	0.478	0.090	0.080	0.486	0.103	0.092	0.502	0.070	0.069	0.544	0.126	0.149	0.528	0.110	0.119
a^*_{62}	0.289	0.310	0.101	0.098	0.315	0.151	0.117	0.335	0.100	0.083	0.359	0.224	0.183	0.353	0.197	0.152
a^*_{63}	0.204	0.235	0.114	0.096	0.247	0.150	0.110	0.270	0.115	0.079	0.205	0.141	0.178	0.179	0.169	0.144
a^*_{64}	0.158	0.144	0.080	0.095	0.138	0.091	0.106	0.175	0.077	0.077	0.094	0.162	0.178	0.131	0.143	0.140
a^*_{65}	0.129	0.162	0.131	0.095	0.149	0.148	0.102	0.183	0.106	0.074	0.119	0.168	0.165	0.096	0.161	0.137
a^*_{66}	0.764	0.715	0.069	0.049	0.727	0.085	0.049	0.773	0.060	0.044	0.702	0.088	0.075	0.712	0.083	0.063

Note: $\theta \equiv$ parameter, DGP $\equiv$ data generating value, $\bar{\theta} \equiv$ average parameter estimate,
MSE $\equiv$ root mean square error, $\overline{\text{PSD}} \equiv$ average posterior standard deviation,
$\overline{\text{ASE}} \equiv$ average asymptotic standard error.

Table 4: Coefficient specification (3.1), Diagonal Σ^* specification

		Bayesian Inference			*MSM-GHK*			*SML-GHK*			*MSM-KS*					
											ρ = .10			ρ = .20		
θ	DGP	$\bar{\bar{\theta}}$	MSE	$\overline{\text{PSD}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$
β^*_{11}	0.000	-0.118	0.142	0.096	-0.021	0.164	0.161	-0.025	0.133	0.117	-0.013	0.109	0.293	0.044	0.347	0.237
β^*_{21}	0.000	-0.162	0.189	0.115	-0.015	0.143	0.154	0.074	0.159	0.150	0.097	0.248	0.261	0.158	0.362	0.220
β^*_{31}	0.000	-0.141	0.173	0.108	-0.010	0.131	0.153	0.122	0.222	0.167	0.020	0.183	0.243	0.099	0.285	0.212
β^*_{41}	0.000	-0.163	0.208	0.113	0.006	0.140	0.123	-0.005	0.156	0.097	-0.035	0.177	0.215	-0.014	0.157	0.166
β^*_{51}	0.000	-0.158	0.178	0.107	-0.032	0.099	0.120	-0.070	0.123	0.107	-0.023	0.181	0.209	-0.037	0.156	0.163
β^*_{61}	0.000	-0.124	0.219	0.108	-0.022	0.154	0.119	-0.101	0.177	0.114	-0.109	0.194	0.191	-0.110	0.176	0.154
β^*_{12}	0.000	0.009	0.030	0.030	0.059	0.169	0.148	0.089	0.151	0.140	0.059	0.260	0.266	0.129	0.355	0.221
β^*_{22}	0.000	-0.010	0.039	0.031	0.008	0.098	0.146	0.106	0.144	0.147	0.028	0.170	0.236	0.071	0.188	0.205
β^*_{32}	0.000	-0.013	0.033	0.030	-0.059	0.140	0.144	0.083	0.178	0.144	-0.017	0.164	0.232	0.007	0.263	0.201
β^*_{42}	0.000	0.004	0.043	0.031	0.005	0.026	0.029	0.004	0.026	0.030	0.009	0.031	0.033	0.009	0.030	0.031
β^*_{52}	0.000	0.007	0.026	0.030	-0.012	0.038	0.029	-0.011	0.036	0.031	-0.012	0.027	0.032	-0.011	0.033	0.030
β^*_{62}	0.000	0.000	0.029	0.030	-0.015	0.035	0.029	-0.017	0.038	0.031	-0.008	0.026	0.032	-0.012	0.032	0.030
γ^*_{11}	0.707	0.763	0.070	0.039	0.704	0.049	0.048	0.725	0.050	0.041	0.728	0.049	0.075	0.771	0.081	0.057
γ^*_{21}	0.707	0.771	0.080	0.082	0.710	0.034	0.076	0.750	0.063	0.069	0.729	0.127	0.129	0.777	0.116	0.098
γ^*_{31}	0.707	0.768	0.083	0.077	0.709	0.058	0.074	0.774	0.092	0.073	0.782	0.122	0.114	0.821	0.131	0.090
γ^*_{41}	0.707	0.784	0.110	0.081	0.722	0.072	0.074	0.788	0.112	0.074	0.764	0.145	0.113	0.805	0.139	0.091
γ^*_{51}	0.707	0.781	0.089	0.076	0.723	0.043	0.074	0.807	0.114	0.077	0.761	0.100	0.109	0.810	0.127	0.088
γ^*_{61}	0.707	0.757	0.099	0.078	0.698	0.082	0.074	0.783	0.112	0.074	0.749	0.126	0.109	0.778	0.113	0.088
γ^*_{71}	-0.707	-0.694	0.040	0.047	-0.715	0.033	0.047	-0.737	0.046	0.042	-0.733	0.046	0.073	-0.776	0.080	0.056
ς^*_1	0.000	0.281	0.327	0.210	0.017	0.061	0.043	0.013	0.056	0.044	0.001	0.042	0.059	0.005	0.040	0.052
ς^*_2	0.000	0.373	0.420	0.243	0.008	0.078	0.122	-0.078	0.148	0.123	-0.039	0.117	0.194	-0.054	0.128	0.159
ς^*_3	0.000	0.293	0.361	0.218	0.025	0.072	0.042	0.047	0.137	0.054	0.023	0.077	0.054	0.024	0.068	0.049
ς^*_4	0.000	0.378	0.465	0.242	0.112	0.218	0.122	0.027	0.273	0.131	0.102	0.314	0.190	0.092	0.247	0.158
ς^*_5	0.000	0.320	0.387	0.226	0.035	0.097	0.042	0.052	0.152	0.055	0.021	0.046	0.053	0.024	0.057	0.049
ς^*_6	0.000	0.307	0.519	0.225	0.073	0.231	0.016	0.095	0.301	0.040	0.066	0.208	0.025	0.065	0.205	0.022

Note: $\theta \equiv$ parameter, DGP $\equiv$ data generating value, $\bar{\bar{\theta}} \equiv$ average parameter estimate,
MSE $\equiv$ root mean square error, $\overline{\text{PSD}} \equiv$ average posterior standard deviation,
$\overline{\text{ASE}} \equiv$ average asymptotic standard error.

are often not much larger than those for Bayesian inference or SML-GHK. MSM-KS (ρ = 0.10) is worst in 3 of 7 cases and MSM-KS (ρ = 0.20) is worst in 4 of 7 cases.

For the β^*_{j2} the effect of imposing the diagonal Σ restriction is inconsistent. For MSM-GHK, as compared to table 1, the MSE rise substantially (50% to 500%) for all the β^*_{j2} except β^*_{42}, for which it falls substantially. SML-GHK and MSM-KS follow similar patterns, but for Bayesian inference the changes in the MSEs form table 1 are negligible. The MSE for Bayesian inference are best in 4 of 6 cases, while MSM-GHK, SML-GHK and MSM-KS (ρ = 0.10) are each best or tied for best in 1 of 6 cases. MSM-KS (ρ = 0.20) is worst in 3 of 6 cases, and Bayesian inference, MSM-GHK and SML-GHK are each worst in 1 of 6 cases. Overall, Bayesian inference has the edge, with MSM-GHK (second best in 3 of 6 cases) second, SML-GHK third and MSM-KS again generally dominated.

For the β^*_{j1} the change in MSEs from table 1 is again inconsistent. For MSM-GHK the MSE rise noticeably in 3 of 6 cases and fall noticeably in 2 cases. For SML-GHK the MSE rise substantially in 5 of 6 cases. For Bayesian inference and MSM-KS (ρ = 0.20) the MSE rise substantially in all 6 cases. For MSM-KS (ρ = 0.10) the MSE fall substantially in 2 cases and rise in 2 cases. In terms of MSEs, MSM-GHK is best in 5 of 6 cases and MSM-KS (ρ = 0.10) is best in the other. MSM-KS (ρ = 0.20) is worst in 3 of 6 cases, Bayesian inference is worst in 2 of 6 cases and MSM-KS (ρ = 0.10) is worst in the other. Overall, MSM-GHK appears best, with other methods difficult to rank.

For the ξ_j, MSM-KS (ρ = 0.20) is best in 3 of 6 cases, MSM-GHK is best in 2 of 6 cases, and MSM-KS (ρ = 0.10) is best in 1 of 6 cases. SML-GHK generally performs worst among the classical methods. The Bayesian posterior means for the ξ_j are all upward biased with posterior standard deviations that are nearly as large as the bias. The numerical standard errors (not reported here) were also quite large—as great as one-half the posterior standard deviation. These results are all consistent with a posterior density that is nearly flat over a wide range of the ξ_j.

Overall, in table 4, MSM-GHK appears to dominate other methods. It clearly has the smallest MSE for the γ^*_{j1}, performs best overall for the β^*_{j1}, is dominated for the β^*_{j2} by Bayesian inference but remains better than other methods, and is second best for the ξ_j. Again, MSM-KS is dominated by other methods.

In table 5 we impose both the restrictions that the γ^*_{j1} are equal for all j and that Σ is diagonal. These restrictions result in some dramatic reductions in MSEs from the unrestricted model (table 1), the model with only the equal γ^*_{j1} restriction imposed (table 2), and the model with only the diagonal Σ restriction imposed (table 4). For example, for MSM-GHK the MSE for γ^*_{11} is 3 to 5.5 times smaller than those for the individual γ^*_{j1} in table 1, 1.5 to 4 times smaller than for the individual γ^*_{j1} in table 4, and 3 times smaller than for γ^*_{11} in table 2. For SML-GHK the MSE is roughly 10 times smaller than those for the individual γ^*_{j1} in table 1, 2 to 4.5 times smaller than for the individual γ^*_{j1} in table 4, and 2 times smaller than for γ^*_{11} in table 2. For Bayesian inference the MSE is roughly 30% to 60% smaller than those for the individual γ^*_{j1} in table 1, 20% to 75% smaller than for the individual γ^*_{j1} in table 4, and 33% smaller than for γ^*_{11} in table 2. Similar substantial improvements are also observed for MSM-KS.

In terms of MSE for γ^*_{11}, MSM-GHK is best at 0.020, SML-GHK is second at 0.026, Bayesian inference is third at 0.029, MSM-KS (ρ = 0.10) is fourth at 0.040, and MSM-KS (ρ = 0.20) is fifth at 0.070.

For the β^*_{j2} Bayesian inference is best or tied for best in 4 of 6 cases, and SML-GHK and MSM-KS ($\rho = 0.10$) are each best in 1 of 6 cases. MSM-KS ($\rho = 0.10$) is worst in 3 of 6 cases, and Bayesian Inference, MSM-GHK and SML-GHK are each worst in 1 of 6 cases. Overall, Bayesian inference has the edge, with MSM-GHK (second best in 4 of 6 cases) second, followed by SML-GHK. Again, MSM-KS is generally dominated by other methods.

For the β^*_{j1} Bayesian inference is best or tied for best in 4 of 6 cases. SML-GHK and MSM-GHK give very similar MSEs. Again, MSM-KS is dominated by other methods.

For the ξ_j, MSM-GHK is best in 5 of 6 cases. SML-GHK and MSM-KS ($\rho = 0.20$) give similar MSEs, and MSM-KS ($\rho = 0.10$) gives the largest MSEs among the classical methods. Bayesian posterior means for the ξ_j again show a substantial upward bias. This bias is smaller than in table 4, but great enough that Bayesian inference has the highest MSE in all 6 cases.

Overall, MSM-GHK performs best in table 5. It gives the smallest MSE for γ^*_{11}, and generally has the smallest MSEs for the ξ_j. For the β^*_{j1} and β^*_{j2} Bayesian inference is best with MSM-GHK edging out SML-GHK for second best. Clearly, if one's primary interest is in the β^*_{jk}, then Bayesian inference dominates other methods.

In table 6 we impose the restrictions that the γ^*_{j1} are equal for all j, that Σ is diagonal, and that the β^*_{jk} are zero for all j and k = 1, 2. Across all methods, this reduces the MSEs for γ^*_{11} by from 20% to 60% from those in table 5. The effect on the MSEs for the ξ_j is mixed for the classical methods, but Bayesian posterior means improve to the point that they show the smallest or tied for smallest MSE in 5 out of 6 cases. The MSE for all 7 free parameters are extremely close for MSM-GHK, SML-GHK and Bayesian Inference. Thus, it is impossible to chose among these methods. MSM-KS is again dominated by other methods.

Table 5: Coefficient specification (3.4), Diagonal Σ^* specification

		Bayesian Inference			*MSM-GHK*			*SML-GHK*			*MSM-KS*					
											$\rho = .10$			$\rho = .20$		
θ	DGP	$\bar{\bar{\theta}}$	MSE	$\overline{\text{PSD}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$
β^*_{11}	0.000	-0.011	0.053	0.049	0.012	0.115	0.081	0.007	0.105	0.074	0.027	0.237	0.111	0.021	0.177	0.099
β^*_{21}	0.000	-0.042	0.052	0.050	0.011	0.085	0.079	0.038	0.106	0.082	0.044	0.204	0.104	0.046	0.147	0.096
β^*_{31}	0.000	-0.030	0.079	0.049	0.012	0.120	0.079	0.045	0.127	0.081	0.025	0.220	0.099	0.039	0.180	0.094
β^*_{41}	0.000	-0.026	0.059	0.050	0.004	0.053	0.051	0.002	0.054	0.047	0.008	0.085	0.067	0.006	0.070	0.059
β^*_{51}	0.000	-0.027	0.073	0.049	-0.020	0.041	0.051	-0.028	0.040	0.050	-0.002	0.067	0.064	-0.015	0.053	0.058
β^*_{61}	0.000	-0.022	0.052	0.049	-0.019	0.067	0.051	-0.038	0.079	0.050	-0.026	0.104	0.063	-0.035	0.087	0.058
β^*_{12}	0.000	0.007	0.030	0.029	0.061	0.131	0.078	0.064	0.115	0.081	0.066	0.218	0.104	0.079	0.188	0.095
β^*_{22}	0.000	-0.009	0.039	0.030	0.024	0.071	0.077	0.048	0.080	0.079	0.027	0.120	0.098	0.045	0.115	0.092
β^*_{32}	0.000	-0.012	0.032	0.030	-0.010	0.071	0.076	0.029	0.077	0.078	-0.002	0.153	0.098	0.005	0.125	0.091
β^*_{42}	0.000	0.004	0.041	0.029	0.006	0.026	0.029	0.003	0.025	0.029	0.009	0.032	0.033	0.008	0.030	0.031
β^*_{52}	0.000	0.005	0.026	0.030	-0.010	0.037	0.029	-0.011	0.036	0.030	-0.012	0.030	0.032	-0.011	0.034	0.030
β^*_{62}	0.000	0.001	0.027	0.030	-0.013	0.033	0.029	-0.016	0.036	0.030	-0.008	0.026	0.032	-0.011	0.030	0.030
γ^*_{11}	0.707	0.721	0.029	0.026	0.710	0.020	0.027	0.722	0.026	0.026	0.722	0.040	0.035	0.770	0.070	0.028
ξ^*_1	0.000	0.061	0.130	0.085	0.001	0.051	0.053	-0.014	0.055	0.052	0.006	0.074	0.068	-0.004	0.066	0.062
ξ^*_2	0.000	0.117	0.159	0.093	0.007	0.042	0.035	0.010	0.050	0.036	0.010	0.036	0.040	0.009	0.043	0.038
ξ^*_3	0.000	0.066	0.123	0.087	-0.005	0.066	0.054	-0.021	0.071	0.053	0.001	0.099	0.065	-0.011	0.083	0.061
ξ^*_4	0.000	0.084	0.116	0.089	0.015	0.043	0.034	0.020	0.057	0.037	0.017	0.065	0.040	0.021	0.061	0.038
ξ^*_5	0.000	0.061	0.143	0.086	0.008	0.041	0.053	-0.009	0.046	0.053	0.009	0.066	0.066	0.004	0.050	0.061
ξ^*_6	0.000	0.059	0.113	0.086	0.024	0.062	0.035	0.029	0.083	0.038	0.032	0.076	0.040	0.030	0.076	0.038

Note: $\theta \equiv$ parameter, DGP $\equiv$ data generating value, $\bar{\bar{\theta}} \equiv$ average parameter estimate,
MSE $\equiv$ root mean square error, $\overline{\text{PSD}} \equiv$ average posterior standard deviation,
$\overline{\text{ASE}} \equiv$ average asymptotic standard error.

Table 6: Coefficient specification (3.5), Diagonal Σ^* specification

		Bayesian Inference			*MSM-GHK*			*SML-GHK*			*MSM-KS*					
											$\rho = .10$			$\rho = .20$		
θ	DGP	$\bar{\bar{\theta}}$	MSE	$\overline{\text{PSD}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$
γ^*_{11}	0.707	0.702	0.015	0.018	0.704	0.012	0.018	0.704	0.013	0.017	0.717	0.020	0.022	0.759	0.054	0.018
ξ^*_1	0.000	0.011	0.064	0.046	0.015	0.074	0.048	0.012	0.071	0.046	0.013	0.106	0.058	0.018	0.102	0.053
ξ^*_2	0.000	0.025	0.054	0.047	0.029	0.066	0.048	0.028	0.061	0.047	0.032	0.088	0.058	0.038	0.087	0.053
ξ^*_3	0.000	-0.004	0.038	0.046	-0.003	0.038	0.047	0.006	0.042	0.046	-0.005	0.070	0.056	0.000	0.055	0.052
ξ^*_4	0.000	0.020	0.051	0.047	0.024	0.051	0.048	0.028	0.051	0.047	0.024	0.056	0.057	0.032	0.061	0.053
ξ^*_5	0.000	-0.009	0.047	0.045	-0.008	0.044	0.047	0.002	0.047	0.046	-0.015	0.070	0.055	-0.005	0.062	0.052
ξ^*_6	0.000	0.002	0.052	0.045	0.003	0.054	0.047	0.016	0.055	0.047	0.006	0.075	0.056	0.012	0.077	0.052

Note: $\theta \equiv$ parameter, DGP $\equiv$ data generating value, $\bar{\bar{\theta}} \equiv$ average parameter estimate,
MSE $\equiv$ root mean square error, $\overline{\text{PSD}} \equiv$ average posterior standard deviation,
$\overline{\text{ASE}} \equiv$ average asymptotic standard error.

In tables 7 to 9 we impose the restriction that Σ is the identity matrix. This restriction does not lead to particularly interesting models, since the whole point of simulation estimation is to make estimation feasible without such strong covariance restrictions. Nevertheless, it is interesting to look at these results to see how attempting to estimate large numbers of covariance matrix parameters reduces the precision with which regressor coefficients are estimated.

In table 7 only the restriction that Σ is diagonal is imposed. This restriction produces some dramatic MSE reductions. For example, as compared to table 1, for MSM-GHK this reduces the MSE for the γ^*_{j1} by factors of 2 to 4.5. As compared to table 4, where only a diagonal Σ restriction is imposed, the MSE reductions range from less than 10% to factors of 3.5. For the β^*_{j2} the MSE reductions from table 1 are only about 5% to 10%, but for the β^*_{j1} the MSE reductions are from 60% to 80%.

In table 7 the MSE for all 18 free parameters are so close across MSM-GHK, SML-GHK and Bayesian inference so as to make it impossible to choose among these methods. MSM-KS is again dominated in terms of MSEs for all parameters. However, MSM-KS ($\rho = 0.10$) clearly dominates MSM-KS ($\rho = 0.20$), and for the γ^*_{j1} its degree of inferiority is not nearly so great as in previous experiments.

In table 8 we impose the restrictions that the γ^*_{j1} are equal for all j and that Σ is the identity matrix. In some cases this leads to dramatic MSE reductions. For example, for MSM-GHK the MSE for γ^*_{11} is roughly 5 to 10 times smaller than those for the individual γ^*_{j1} in table 1. It is 6 times smaller than that for γ^*_{11} in table 2, but not quite 2 times smaller than that for γ^*_{11} in table 5, where Σ is only restricted to be diagonal. Thus, comparing tables 2, 5, and 8, we see that most of the reduction in MSE for γ^*_{11} is achieved by imposing diagonality on Σ rather than going all the way to an identity matrix restriction. For the β^*_{j1}, the MSEs for MSM-

Table 7: Coefficient specification (3.1), Scalar Σ^* specification

		Bayesian Inference			*MSM-GHK*			*SML-GHK*			*MSM-KS*					
											ρ = .10			ρ = .20		
θ	DGP	$\bar{\bar{\theta}}$	MSE	$\overline{\text{PSD}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$
β^*_{11}	0.000	0.010	0.050	0.036	0.009	0.049	0.036	0.008	0.047	0.035	0.006	0.053	0.040	0.008	0.053	0.036
β^*_{21}	0.000	0.017	0.041	0.035	0.017	0.041	0.036	0.013	0.040	0.035	0.021	0.045	0.041	0.021	0.046	0.036
β^*_{31}	0.000	-0.002	0.028	0.035	-0.003	0.026	0.036	-0.004	0.025	0.035	-0.014	0.035	0.039	-0.010	0.032	0.035
β^*_{41}	0.000	0.014	0.036	0.035	0.015	0.035	0.036	0.010	0.031	0.035	0.015	0.030	0.041	0.015	0.033	0.036
β^*_{51}	0.000	-0.008	0.032	0.036	-0.008	0.031	0.035	-0.013	0.033	0.036	-0.016	0.043	0.039	-0.011	0.038	0.035
β^*_{61}	0.000	0.005	0.041	0.035	0.004	0.040	0.036	0.001	0.039	0.035	0.000	0.044	0.039	0.005	0.047	0.036
β^*_{12}	0.000	0.009	0.029	0.029	0.007	0.028	0.029	0.006	0.027	0.029	0.011	0.034	0.033	0.011	0.033	0.031
β^*_{22}	0.000	-0.008	0.037	0.029	-0.008	0.039	0.029	-0.008	0.037	0.029	-0.009	0.033	0.032	-0.007	0.037	0.030
β^*_{32}	0.000	-0.010	0.030	0.029	-0.010	0.030	0.029	-0.011	0.030	0.029	-0.008	0.024	0.032	-0.009	0.027	0.030
β^*_{42}	0.000	0.004	0.040	0.029	0.003	0.039	0.029	0.002	0.038	0.029	0.011	0.037	0.032	0.006	0.040	0.031
β^*_{52}	0.000	0.006	0.025	0.029	0.006	0.024	0.029	0.006	0.023	0.029	0.001	0.031	0.032	0.005	0.028	0.030
β^*_{62}	0.000	0.002	0.026	0.029	0.003	0.026	0.029	0.001	0.027	0.029	0.008	0.027	0.032	0.005	0.027	0.030
γ^*_{11}	0.707	0.707	0.027	0.027	0.707	0.027	0.028	0.701	0.027	0.027	0.722	0.032	0.031	0.762	0.063	0.027
γ^*_{21}	0.707	0.689	0.027	0.027	0.688	0.028	0.028	0.684	0.030	0.027	0.698	0.025	0.032	0.739	0.039	0.028
γ^*_{31}	0.707	0.707	0.032	0.027	0.707	0.031	0.027	0.704	0.032	0.027	0.727	0.037	0.029	0.763	0.065	0.027
γ^*_{41}	0.707	0.700	0.022	0.027	0.701	0.022	0.028	0.697	0.023	0.027	0.711	0.029	0.032	0.751	0.050	0.028
γ^*_{51}	0.707	0.714	0.028	0.028	0.715	0.028	0.027	0.712	0.028	0.027	0.731	0.044	0.030	0.768	0.068	0.027
γ^*_{61}	0.707	0.703	0.024	0.027	0.703	0.024	0.027	0.699	0.026	0.027	0.720	0.039	0.029	0.756	0.057	0.027
γ^*_{71}	-0.707	-0.710	0.032	0.027	-0.709	0.031	0.027	-0.702	0.030	0.027	-0.716	0.034	0.031	-0.760	0.062	0.027

Note: $\theta \equiv$ parameter, DGP $\equiv$ data generating value, $\bar{\bar{\theta}} \equiv$ average parameter estimate,
MSE $\equiv$ root mean square error, $\overline{\text{PSD}} \equiv$ average posterior standard deviation,
$\overline{\text{ASE}} \equiv$ average asymptotic standard error.

Table 8: Coefficient specification (3.4), Scalar Σ^* specification

		Bayesian Inference			*MSM-GHK*			*SML-GHK*			*MSM-KS*					
											ρ = .10			ρ = .20		
θ	DGP	$\bar{\bar{\theta}}$	MSE	$\overline{\text{PSD}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$
β^*_{11}	0.000	0.009	0.036	0.029	0.008	0.035	0.029	0.007	0.033	0.029	0.009	0.039	0.032	0.009	0.040	0.030
β^*_{21}	0.000	0.006	0.023	0.029	0.006	0.024	0.029	0.003	0.024	0.028	0.012	0.028	0.032	0.009	0.027	0.031
β^*_{31}	0.000	-0.004	0.034	0.029	-0.003	0.033	0.029	-0.004	0.032	0.029	-0.010	0.032	0.032	-0.008	0.034	0.030
β^*_{41}	0.000	0.009	0.035	0.029	0.010	0.034	0.029	0.007	0.033	0.028	0.011	0.031	0.032	0.011	0.034	0.031
β^*_{51}	0.000	-0.006	0.026	0.029	-0.005	0.024	0.029	-0.008	0.026	0.029	-0.009	0.028	0.032	-0.006	0.028	0.030
β^*_{61}	0.000	0.001	0.029	0.029	0.001	0.028	0.029	-0.001	0.026	0.029	0.002	0.030	0.032	0.003	0.032	0.030
β^*_{12}	0.000	0.008	0.029	0.029	0.008	0.029	0.029	0.006	0.027	0.029	0.011	0.034	0.033	0.011	0.033	0.031
β^*_{22}	0.000	-0.008	0.038	0.029	-0.007	0.039	0.029	-0.008	0.037	0.029	-0.009	0.034	0.032	-0.007	0.038	0.030
β^*_{32}	0.000	-0.010	0.031	0.029	-0.009	0.030	0.029	-0.011	0.031	0.029	-0.008	0.025	0.032	-0.009	0.028	0.030
β^*_{42}	0.000	0.003	0.040	0.029	0.003	0.039	0.029	0.002	0.038	0.029	0.011	0.037	0.032	0.006	0.040	0.031
β^*_{52}	0.000	0.006	0.025	0.029	0.006	0.024	0.029	0.005	0.023	0.029	0.001	0.031	0.032	0.005	0.028	0.030
β^*_{62}	0.000	0.002	0.026	0.029	0.003	0.026	0.029	0.001	0.027	0.029	0.007	0.027	0.032	0.005	0.027	0.030
γ^*_{11}	0.707	0.703	0.011	0.011	0.703	0.011	0.012	0.699	0.013	0.011	0.717	0.017	0.013	0.756	0.051	0.011

Note: $\theta \equiv$ parameter, DGP $\equiv$ data generating value, $\bar{\bar{\theta}} \equiv$ average parameter estimate,
MSE $\equiv$ root mean square error, $\overline{\text{PSD}} \equiv$ average posterior standard deviation,
$\overline{\text{ASE}} \equiv$ average asymptotic standard error.

Table 9: Coefficient specification (3.5), Scalar Σ^* specification

		Bayesian Inference			*MSM-GHK*			*SML-GHK*			*MSM-KS*					
											$\rho = .10$			$\rho = .20$		
θ	DGP	$\bar{\bar{\theta}}$	MSE	$\overline{\text{PSD}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$
γ^*_{11}	0.707	0.702	0.012	0.011	0.703	0.011	0.012	0.699	0.013	0.011	0.716	0.016	0.013	0.755	0.050	0.011

Note: $\theta \equiv$ parameter, DGP $\equiv$ data generating value, $\bar{\bar{\theta}} \equiv$ average parameter estimate,
MSE $\equiv$ root mean square error, $\overline{\text{PSD}} \equiv$ average posterior standard deviation,
$\overline{\text{ASE}} \equiv$ average asymptotic standard error.

GHK in table 8 are on the order of 2 times smaller than those in table 2, but the MSE for the β^*_{j2} are almost identical.

In table 8 the MSE for all 13 parameters are almost identical for MSM-GHK, SML-GHK, and Bayesian inference. MSM-KS ($\rho = 0.10$) also produces very similar MSEs for the β^*_{jk}, but its MSE for γ^*_{11} is roughly 50% larger. MSM-KS ($\rho = 0.20$) also produces very similar MSEs for the β^*_{jk}, but it is clearly inferior to other methods for γ^*_{11}.

In table 9 we impose the 3 restrictions that the γ^*_{j1} are equal for all j, that Σ is the identity matrix, and that the β^*_{jk} are zero for all j and k = 1, 2. Thus γ^*_{11} is the only free parameter. MSM-GHK, SML-GHK and SML-GHK produce very similar results (with a slight edge to MSM-GHK) and MSM-KS is dominated by other methods.

Throughout tables 1–9 a number of very clear patterns emerge. One is that the diagonal Σ specification leads to important gains in efficiency. For example, comparing tables 2, 5, and 8 we see that the MSE for γ^*_{11} estimated by MSM-GHK falls from 0.064 to 0.020 with the diagonal Σ restriction imposed, and then falls to 0.011 when the further restriction that Σ is the identity matrix is imposed. Thus, most of the gain that can be achieved by restricting Σ is already achieved by restricting it to be diagonal rather than going all the way to a $\Sigma = I$ restriction. This fact is important, because in many choice applications in areas like marketing and transportation it is a truism among practitioners that the most important deviations from IIA are due to unequal error variances.

Another important pattern is that the MSE for the β^*_{j2} do not fall appreciably when restrictions are placed on the model. For example, comparing tables 2 and 8, we see that for MSM-GHK the MSE for the β^*_{j2} are identical to 3 decimal places in 3 of 6 cases, and differ by less than 4% in the other 3 cases. Meanwhile, the MSE for the β^*_{j1} fall by factors of roughly

2 and the MSE for γ^*_{11} falls by a factor of 6. Comparing tables 1 and 8, we see that the MSE for the β^*_{j2} fall by at most 8%, while the MSE for the β^*_{j1} fall by factors of roughly 4 to 7. The intuitive reason for this pattern is that the β^*_{j2} are identified from household differences in choice probabilities that arise due to differences in the x^*_i, while the parameters β^*_{j1}, γ^*_{j1}, and a_{ij} are all identified primarily from differences in aggregate choice probabilities for alternatives with different attributes. Thus, it is difficult to disentangle differences among the β^*_{j1}, γ^*_{j1}, and a_{ij} because different patterns for these parameters can generate roughly equivalent aggregate choice probabilities conditional on the z^*_{ij}. Thus restrictions on the β^*_{j1}, γ^*_{j1}, and a_{ij} can be expected to lead to substantial MSE reductions among the β^*_{j1}, γ^*_{j1}, and a_{ij} parameters, but not for the β^*_{j2}. (Of course, it is simple to show that so long as x^*_i varies across i, it is impossible to chose alternative values for β^*_{j1}, β^*_{j2}, γ^*_{j1}, and a_{ij} that will generate identical choice probabilities for all i. However, the fact that it is possible to chose alternative values for the model parameters that generate nearly identical choice probabilities is exactly the "fragile identification problem" in the multinomial probit model described by Keane 1992.)

Another pattern is that means of the point estimates and posterior means tend to be close to each other, relative to their distance from the data generating values: when the mean point estimate or mean of the posterior mean across replications is greater than the data generating value for one method, it tends to be greater for all methods. This feature is present for all models and groups of parameters, but is more prevalent for the coefficients than for the variance parameters, and is most striking in the models with smaller numbers of parameters.

In terms of an overall ranking of methods, it is clear that this differs across models. In the unrestricted model of table 1, Bayesian inference has a slight edge over MSM-GHK because it produces slightly lower MSEs for the β^*_{j1} and γ^*_{j1}. SML-GHK performs worse because it has high MSEs for the γ^*_{j1} and the diagonal elements of the a_{ij}, and also because it generates mean ASE that severely understate the MSE. However, as soon as we impose an equality restriction on the γ^*_{j1} in table 2, the advantage of MSM-GHK over SML-GHK disappears, while Bayesian inference appears to have a slight edge in terms of MSE. Performance of all three methods is very similar in table 3, where an equality restriction is also imposed on the β^*_{jk}. In tables 4 and 5, where a diagonal structure is imposed on Σ (and where γ^*_{j1} are unrestricted or forced to be equal, respectively), MSM-GHK dominates other methods. In tables 6 through 9, which are highly restricted models, the MSE for Bayesian inference, MSM-GHK and SML-GHK are essentially indistinguishable.

A clear pattern however, is that MSM-KS is dominated by other methods in all models. The MSEs for this method are consistently greater than for other methods, both when $\rho = 0.10$ and $\rho = 0.20$. For $\rho = 0.20$ asymptotic standard errors are smaller than for $\rho = 0.10$, which is characteristic of kernel smoothers but introduces an important downward bias in test size. MSM-KS also suffers form several other disadvantages. Most important, it requires care in choice of a tuning parameter while the other methods do not. In experiments not reported here in tabular form, we found that convergence required $\rho \geq 0.05$, and for values of ρ above 0.20 bias increased very quickly. Systematic evidence of the increased bias that results from increasing ρ from 0.1 to 0.2 can be seen in tables 1-9.

Finally, we attempted to repeat the experiment using only the first 1,000 observations of the artificial data set. For all of the methods, computational problems emerged for some of the models. As one would expect, problems were concentrated in the models with more free parameters. The most common problems were estimated singularity of the Hessian in the classical methods, and estimated singularity of the variance matrix Σ^* for the Bayesian method.

But across the nine models, computations were carried to completion successfully more often than not. The difficulties involved in obtaining estimates of the larger models using only 1,000 observations are attributable to the relative lack of information contained in discrete choice data that we referred to earlier.

B. Experiment Two

In the second experiment only the model with the γ^*_{j1} restricted to be equal was estimated. This model was estimated using 50 artificial data sets constructed to have properties similar to the Nielsen data on household ketchup purchases. The construction of this data is described in Section 2. In this experiment we compare three estimators of this model:

1. Posterior means using the Gibbs sampling-data augmentation algorithm with m = 10,000 iterations.
2. Method of simulated moments using the GHK probability simulator with M = 30 draws to simulate the choice probabilities and the derivatives that enter the optimal weights, and using Gauss-Newton iterations to solve the simulated moment conditions.
3. Simulated maximum likelihood using the GHK probability simulator with M = 30 draws to simulate the choice probabilities, and using BHHH iterations to maximize the simulated log-likelihood function.

We do not consider MSM-KS because it is already clear from experiment 1 that it is dominated by other methods. There are two reasons why we only consider the model with an equality restriction imposed on the γ^*_{j1} in this experiment. First, we would like to do one Monte-Carlo experiment that is more thorough by using 50 artificial data sets rather than only

10. However, it is not computationally for us to do this many replications on all 9 models. Since, for reasons described earlier, the model with the equality restriction imposed on the γ^*_{j1} is the most realistic model, its seems natural to choose that one. Second, given that the data are less behaved here than in experiment 1, we do not feel it is feasible to estimate the unrestricted model using only 5,000 observations, as severe problems would probably arise in identifying the covariance matrix parameters.

The results of the second experiment are reported in table 10. In terms of MSE, the results for γ^*_{11} are very similar across the three methods. For the β^*_{j2}, Bayesian inference produces MSEs that are usually about 10% to 30% smaller than the classical methods, and it is best or tied for best in 5 of 6 cases. The MSE for MSM-GHK and SML-GHK are very similar. For the β^*_{j1}, Bayesian inference produces MSEs that are often 10% to 50% smaller than the classical methods, and again it is best or tied for best in 5 of 6 cases. Again, MSM-GHK and SML-GHK produce very similar MSEs. For the a_{ij} it is not nearly so obvious how to rank the estimators. Bayesian Inference is best in terms of MSE in 10 of 20 cases, while MSM-GHK is best in 6 of 20 and SML-GHK is best in 4 of 20. Although these figures might appear to give Bayesian Inference a slight edge, it is again true that the MSE for the classical methods are very close, so that in all 10 cases where Bayesian inference is not best it is in fact the worst of the three methods.

One obvious feature of the results in table 10 is that the MSE for all parameters except the β^*_{j2} are generally much larger than that for the corresponding parameter in table 2. This illustrates how the information in the sample that is useful for identifying the parameters β^*_{j1}, γ^*_{j1} and a_{ij} declines when the z^*_{ij} are not IID. On the other hand, there is no general tendency for the MSE for the β^*_{j2} to be larger or smaller in table 10 than in table 2.

Table 10: Coefficient specification (3.2), Unrestricted Σ^* specification

		Bayesian Inference			*MSM-GHK*			*SML-GHK*		
θ	DGP	$\bar{\bar{\theta}}$	MSE	$\overline{\text{PSD}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$	$\bar{\bar{\theta}}$	MSE	$\overline{\text{ASE}}$
β^*_{11}	-0.307	-0.315	0.091	0.102	-0.318	0.106	0.111	-0.319	0.110	0.091
β^*_{21}	-0.961	-0.995	0.186	0.189	-1.066	0.225	0.213	-1.071	0.228	0.182
β^*_{31}	0.163	0.041	0.214	0.173	-0.027	0.282	0.219	-0.034	0.289	0.183
β^*_{41}	-0.946	-1.042	0.535	0.381	-1.593	1.042	0.818	-1.591	1.044	0.591
β^*_{51}	1.402	1.101	0.423	0.237	1.230	0.307	0.298	1.226	0.308	0.240
β^*_{61}	0.954	0.921	0.102	0.105	0.891	0.116	0.108	0.888	0.118	0.097
β^*_{12}	-0.033	-0.031	0.016	0.018	-0.029	0.018	0.021	-0.029	0.018	0.018
β^*_{22}	-0.011	-0.019	0.029	0.028	-0.009	0.032	0.033	-0.010	0.032	0.028
β^*_{32}	-0.040	-0.034	0.028	0.029	-0.019	0.036	0.034	-0.019	0.035	0.028
β^*_{42}	-0.035	-0.035	0.046	0.043	0.001	0.069	0.073	0.000	0.068	0.047
β^*_{52}	-0.359	-0.430	0.116	0.077	-0.397	0.098	0.099	-0.398	0.097	0.088
β^*_{62}	-0.171	-0.171	0.022	0.026	-0.170	0.022	0.027	-0.171	0.022	0.025
γ^*	-1.981	-1.958	0.121	0.120	-1.991	0.122	0.125	-1.997	0.122	0.118
a^*_{11}	1.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.000	0.000
a^*_{21}	0.615	0.569	0.210	0.189	0.557	0.221	0.196	0.541	0.214	0.143
a^*_{22}	1.019	1.061	0.139	0.129	1.031	0.127	0.129	1.032	0.125	0.115
a^*_{31}	0.410	0.386	0.275	0.243	0.448	0.205	0.265	0.457	0.195	0.147
a^*_{32}	0.443	0.358	0.358	0.325	0.346	0.255	0.401	0.339	0.256	0.163
a^*_{33}	1.407	1.347	0.213	0.171	1.305	0.185	0.184	1.299	0.189	0.150
a^*_{41}	0.322	0.366	0.435	0.384	0.514	0.451	0.654	0.530	0.453	0.244
a^*_{42}	0.401	0.380	0.391	0.429	0.339	0.424	1.061	0.324	0.420	0.254
a^*_{43}	1.483	1.275	0.458	0.347	0.906	0.672	0.808	0.849	0.729	0.268
a^*_{44}	1.096	0.991	0.216	0.249	1.460	0.550	0.477	1.457	0.551	0.313
a^*_{51}	0.351	0.290	0.273	0.291	0.285	0.237	0.333	-0.293	0.229	0.190
a^*_{52}	0.024	0.027	0.430	0.370	0.104	0.328	0.655	0.107	0.337	0.200
a^*_{53}	0.497	0.253	0.471	0.357	0.255	0.399	0.623	0.238	0.406	0.202
a^*_{54}	0.238	0.189	0.312	0.362	0.103	0.277	1.461	0.082	0.286	0.187
a^*_{55}	0.905	0.859	0.208	0.234	1.003	0.283	0.428	1.004	0.284	0.249
a^*_{61}	0.506	0.514	0.147	0.165	0.524	0.154	0.162	0.510	0.166	0.113
a^*_{62}	0.641	0.603	0.231	0.223	0.407	0.333	0.245	0.368	0.363	0.137
a^*_{63}	0.955	0.970	0.143	0.170	0.760	0.251	0.185	0.726	0.282	0.130
a^*_{64}	0.361	0.387	0.184	0.216	0.223	0.221	0.376	0.197	0.235	0.126
a^*_{65}	-0.100	0.036	0.256	0.250	0.059	0.200	0.584	0.053	0.194	0.121
a^*_{66}	0.845	0.618	0.240	0.142	1.025	0.217	0.146	1.062	0.247	0.101

Note: $\theta \equiv$ parameter, DGP $\equiv$ data generating value, $\bar{\bar{\theta}} \equiv$ average parameter estimate, MSE $\equiv$ root mean square error, $\overline{\text{PSD}} \equiv$ average posterior standard deviation, $\overline{\text{ASE}} \equiv$ average asymptotic standard error.

Another obvious feature of table 10 is that large biases are apparent for several of the parameters. Examples of parameters for which highly significant biases appear for one or more of the methods (in the sense that the estimated bias exceeds the empirical standard deviation of the bias by several standard deviations) are β^*_{31}, β^*_{41}, β^*_{51}, β^*_{61}, a_{43}, a_{44}, a_{53}, a_{63}, a_{64}, and a_{66}. Such substantial biases as appear for these parameters were not apparent for methods other than MSM-KS in table 2. It is also important to note that, when one method shows substantial bias for a particular parameter, in most cases the other methods tend to be biased in the same direction. This evidence indicates that the sample size of 5,000 is not adequate to eliminate small sample bias given the configuration of regressors and parameter values in the second experiment. The fact that a sample size of 5,000 did appear sufficient to render negligible any small sample bias in the first experiment is likely due to the IID property of the regressors in that experiment.

The agreement between empirical MSEs and the means of the asymptotic standard errors or posterior standard deviations is not so close as in the first experiment. For SML-GHK the mean ASE is less than the empirical MSE for 32 of 33 parameters, and the differences are often large. For MSM-GHK the relation between the MSE and ASE depends on the set of parameters considered. For the β^*_{j1}, β^*_{j2}, and γ^*_{j1} there is generally close agreement between MSEs and mean ASEs, and there is no systematic tendency for one to be greater than the other. But for the a_{ij}, the mean ASE overstates the MSE in 13 of 20 cases. More importantly, the degree by which the ASE exceed the MSE is often very substantial. For Bayesian inference, the mean posterior standard deviation lies substantially below the MSE for a few of the β^*_{jk}, particularly for β^*_{41}, β^*_{51}, and β^*_{51}. But for the remaining β^*_{jk}, and also for γ^*_{11}, there is generally close agreement between MSEs and mean PSDs. For the a_{ij}, the agreement between the MSEs and the mean PSDs is better than for the classical methods, and there is no systematic tendency for one to lie above the other (i.e., the mean PSD is less than the MSE in 11 of 20 cases). Also, for Bayesian inference the agreement between the MSE and mean PSD in table 10 generally appears to be just as close as in table 2. This is of course not surprising, because the PSD for Bayesian inference are based on the exact finite sample posterior distributions of parameter value draws rather than on asymptotic sampling theory.

A problem with MSM-GHK that is not apparent from table 10 is the difficulty involved in solving the simulated moment conditions. The Gauss-Newton iterations used to solve the simulated moment conditions are given by:

$$\text{Delta}(\theta) = \sum_{i=1}^{N}\sum_{j=1}^{J} W_{ij}[d_{ij} - \hat{P}(j\,|\,\theta)][\sum_{i=1}^{N}\sum_{j=1}^{J} W_{ij}\partial\hat{P}(\cdot)/\partial\theta]^{-1}$$

where the W_{ij} are simulations of the quantities $\partial\hat{P}(\cdot)/\partial\theta)/\hat{P}(\cdot)$. Recall that these are the asymptotically optimal instruments as $M \rightarrow \infty$, provided that the W_{ij} are evaluated at an initial consistent (but not efficient) estimate of θ. It is also important to note that the matrix $[\sum_{i=1}^{N}\sum_{j=1}^{J} W_{ij}\partial\hat{P}(\cdot)/\partial\theta]^{-1}$ may be replaced by any positive definite matrix and the search algorithm is still guaranteed to eventually find a root. This means, for example, that one may evaluate $[\sum_{i=1}^{N}\sum_{j=1}^{J} W_{ij}\partial\hat{P}(\cdot)/\partial\theta]^{-1}$ at an initial value of θ and then hold it fixed throughout the search. This leads to important gains in computational time because the cost of evaluating derivatives on each iteration is avoided. Note that the calculation of the necessary derivatives for MSM estimation is much more time consuming than for SML estimation. SML requires only that the derivatives of the choice probabilities for the chosen alternatives be calculated, while MSM requires that the derivatives of the choice probabilities for all alternatives be calculated.

One important problem that arises in solving the simulated moment conditions is sensitivity of the results to the construction of the initial consistent estimate θ_0 of θ that is used to form the optimal initial weights and as the starting point for the search algorithm. We find that in models with small numbers of parameters the choice of θ_0 has very little impact on parameter estimates and standard errors, so that, as a practical matter, obtaining an initial consistent estimate of θ is not even necessary. This also tends to be true in larger models if the data is very well behaved (as in experiment 1). Rather than obtaining an initial consistent estimate of θ it is typically sufficient to start at some neutral values (such as covariance matrix set to the identity, regressor coefficients set to zero, and constants set to zero or perhaps to values that replicate the aggregate choice frequencies). But in more difficult estimation contexts (such as experiment 2) the choice of θ_0 becomes important.

Another important problem that arises in solving the simulated moment conditions involves the forming of the derivatives $(\partial\hat{P}(\cdot)/\partial\theta)/\hat{P}(\cdot)$ that are needed to construct the step. In larger models with badly behaved data the Gauss-Newton search algorithm described above, where the derivatives $(\partial\hat{P}(\cdot)/\partial\theta)/\hat{P}(\cdot)$ are held fixed at their initial values, works poorly. We suspect this is because as one moves away from the initial θ the derivatives evaluated at that θ become less useful in determining an optimal step.

These problems of sensitivity of MSM estimates to initial values and difficulty in finding steps once one moves far away from the initial θ were quite severe in the second experiment. We found that very different results were obtained if we started from SML-GHK parameter estimates vs. starting from neutral parameter values, and that the former results were much more reasonable. Thus, we were forced to use SML-GHK estimates as starting values. Obviously, it reduces the appeal of MSM-GHK if one must obtain SML-GHK estimates first.

Overall, in the second experiment, Bayesian inference appears to have a clear advantage over other methods. It dominates in terms of MSE for the β^*_{jk}, and produces MSE that are similar to MSM-GHK and SML-GHK for the other parameters. Also, the agreement between MSE and mean PSD for Bayesian inference is generally better than that between MSE and mean ASE for the classical methods. MSM-GHK and SML-GHK results are similar in terms of MSEs, but MSM-GHK dominates in terms of agreement between the MSE and the mean ASE. It must be remembered, however, that SML-GHK has an important ease of use advantage, since it was necessary to use SML-GHK estimates as starting values for MSM-GHK.

7. Conclusion

Overall, our results indicate that the performance of all the alternative simulation based approaches to inference in the multinomial probit model is satisfactory. With few exceptions, the methods produce parameter estimates or posterior means that are reasonably close to the data generating values even in models with large numbers of parameters.

There are, however, some clear differences in the performance of the methods. Overall, across experiments 1 and 2, it appears that Bayesian inference based on the Gibbs sampler outperforms classical methods. We conclude this for two reasons. First, the MSE of the posterior means generated by Gibbs sampling are usually at least slightly smaller that the MSE of the classical point estimates (although this is certainly not universally true across all models and all sets of parameters). Secondly, the performance of Bayesian inference does not deteriorate so clearly as that of the classical methods when we move from the well behaved data of experiment 1 to the less well behaved data of experiment 2.

Among the classical methods it is difficult to chose between MSM-GHK and SML-GHK. In the completely unrestricted model of table 1, MSM-GHK dominates both in terms of MSE and agreement between mean ASE and MSE (the ASE for SML-GHK tends to underestimate the MSE). But, as soon as equality restrictions are placed on the γ_{j1}^{*} (tables 2 and 3) the performance of the two methods in terms of MSE becomes difficult to distinguish. However, in models where reasonable covariance matrix restrictions are imposed (tables 4 and 5), MSM-GHK has an edge over SML-GHK. In very simple models (tables 6-9) these methods are indistinguishable. In the second experiment, MSM-GHK and SML-GHK produce very similar results in terms of MSE, but MSM-GHK dominates in terms of agreement between MSE and mean ASE. However, MSM-GHK suffers from a clear ease of use disadvantage that arises because of the difficulty in solving the simulated moment conditions when the data is ill behaved as in experiment 2. This actually forces us to use SML-GHK estimates as starting values for MSM-GHK in order to obtain reasonable results. One clear result is that MSM-KS is dominated by all other methods. Thus we conclude that, for models of this type, the choice between estimation methods (i.e., MSM vs. SML) is of secondary importance relative to the choice of probability simulator (i.e., GHK vs. KS vs. other alternatives).

Three of the methods we have studied—MSM-GHK, SML-GHK, and Bayesian inference based on the Gibbs sampling-data augmentation algorithm—provide similar point estimates or posterior means of parameters and similar bases for inference via their asymptotic standard errors or posterior standard deviations in many of the models we consider. In this context, ease of use comparisons and comparisons of computation times become important. For this reason we report in table 11 the results for each method of inference of regressions of computation times from experiment one on the number of parameters in the particular model and this quantity

Table 11: Time Comparison

Method	Constant	P	P^2	Predicted Time		
				P=5	P=20	P=40
Simulated Maximum Likelihood (SML-GHK)	4563.87	452.88	26.95	7502.05	24401.76	65800.15
Method of Simulated Moments (MSM-GHK)	1892.16	1374.02	7.08	8939.19	32203.37	68176.12
Kernel-Smoothing (MSM-KS)	3364.89	181.86	18.23	4729.95	14294.10	39807.16
Bayesian Inference *	11760.23	242.34	-4.46	12860.41	14822.49	14315.53
Bayesian Inference **	658.03	785.31	-4.77	4465.32	14455.65	24435.83

* 10,000 iterations

** level of relative numerical efficiency fixed so that ratio of numerical variance to posterior variance is 0.10 for γ^*_{11}

Note: $P \equiv$ number of parameters in the model. Time reported in CPU seconds on a SUN Sparc 10/51.

squared. We also report the predicted computation times for models with 5, 20, and 40 parameters. Clearly, the computation times for all the methods are of the same order of magnitude. Computation times for SML-GHK and MSM-GHK are very similar in the first experiment, but MSM-KS dominates by roughly a factor of 1.5 to 2. This fact is partially explainable by the fact that we had difficulties in achieving convergence of the MSM-KS parameter estimates to the same tolerance level as was obtained for SML-GHL and MSM-GHK. In a number of the runs in experiment one, the MSM-KS algorithm terminated due to inability to find an objective function improving step, rather than because convergence to the desired tolerance had been achieved. Thus, the faster timings for MSM-KS are partially due to early termination of the search algorithm. Bayesian inference using m = 10,000 produces similar times across models regardless of the number of parameters. This is because the main time cost in the Gibbs sampling algorithm is in drawing the latent variables, and this cost is independent of the number of parameters. For the classical methods the main component of cost is the calculation of derivatives, and this grows linearly with the number of parameters. Thus, Gibbs sampling has a timing disadvantage relative to the classical methods for small models, but a timing advantage in larger models. We also determined the number of iterations, m, needed to achieve a relative numerical efficiency level (defined as the ratio of numerical variance to posterior variance) of 10% for the parameter γ_{11}^{*} in each of the nine models of experiment one, and calculated the time necessary for Bayesian inference at those levels of m. In that case, the time requirement for Bayesian inference does rise substantially with the number of parameters. Given the fixed level of relative numerical efficiency, Bayesian inference still maintains a time advantage over the classical methods for large models, and it no longer suffers a clear disadvantage for small models.

The differences in computation times that appear in table 11 do not seem great enough to dictate choice of method. Note that computation times will depend on the parameters M for the probability simulators and m for the Gibbs-sampling data augmentation algorithm. It appears that our choices of M = 30 for MSM-GHK and SML-GHK, M = 100 for MSM-KS, and m = 10,000 for Gibbs sampling generate roughly comparable computation times for all three methods (i.e., well within an order of magnitude). Other reasonable choices for M and m would not drastically alter this conclusion. Given this, we feel that programming time and ease of use should be dominant considerations. Here, Bayesian inference and SML appear to have the edge over MSM for the reasons discussed in section 6B.

It is impossible, based on the results presented here, to provide any universal comparison of computation times for the different methods, or any comparison across methods of accuracy that can be achieved per hour of computation. Furthermore, extrapolation of the computation times given here to other contexts may be very misleading. There are two main reasons for this. First, the most important components of time for any method are programming time and actual program "babysitting" time (i.e., the time one must devote to obtaining starting values, restarting programs that have "bombed" for various reasons, tinkering with optimization algorithms in order to achieve convergence to an optimum, etc.). The real cost of these types of time inputs is far greater than the cost of CPU time, yet these types of costs are extremely difficult to quantify. Second, there are a myriad of reasons why even purely computational times are very difficult to quantify exactly. As an example, note that MSM estimation requires the initial calculation of the derivatives of the probabilities of all possible choices in order to construct the optimal weighting matrix. But on subsequent iterations this matrix is held fixed and derivative calculations are not necessary. On the other hand, SML requires the calculation of derivatives on each iteration, but only for the chosen alternatives. Thus, the calculation of the initial weighting matrix for MSM is very time consuming, while subsequent iterations are relatively fast. For SML all iterations are roughly equally time consuming. Clearly, there is some number of iterations large enough that MSM is faster than SML, but how many iterations will be necessary in order to achieve convergence to an optimum is completely problem specific, both in terms of model and data set. Also, MSM can become very slow if a number of restarts, with recalculation of the weighting matrix, are necessary to achieve convergence. For Bayesian inference based on Gibbs sampling on the other hand, the number of cycles necessary to achieve a given level of numerical accuracy will depend on the serial correlation in the Gibbs draws, which is again completely problem specific. Given these three caveats, we are unable to provide any universal time comparisons that go beyond the simple statement of how long inference required in our *specific* problems. Nevertheless, our experience leads us to believe that implementation of all the methods that we have considered in workstation environments will be comparable and quite feasible for a wide range of problems.

We conclude by cautioning that our results on relative performance of methods are specific to the cross sectional multinomial probit model. For instance, in our other work (Geweke, Keane, and Runkle 1994) we find clear advantages of MSM-GHK over SML-GHK in the context of certain types of panel data probit models.

References

Albert, J., and Chib, S. 1992. Bayesian analysis of binary and polychotomous data. *Journal of the American Statistical Association,* forthcoming.

Albright, R.; Lerman, S.; and Manski, C. 1977. Report on the development of an estimation program for the multinomial probit model. Report prepared by Cambridge Systematics for the Federal Highway Administration.

Borsch-Supan, A., and Hajivassiliou, V. 1993. Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models. *Journal of Econometrics* 58: 347-68.

Bunch, D. S. 1991. Estimability in the multinomial probit model. *Transportation Research B,* 25: 1–12.

Dansie, B. 1985. Parameter estimability in the multinomial probit model. *Transportation Research B,* 19: 526–28.

Elrod, T., and Keane, M. 1994. A factor-analytic probit model for estimating market structure in panel data. *Journal of Marketing Research*, forthcoming.

Gelfand, A. E., and Smith, A. F. M. 1990. Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85: 398–409.

Geweke, J. 1986. Exact inference in the inequality constrained normal linear regression model. *Journal of Applied Econometrics* 1: 127–42.

__________. 1988. Antithetic acceleration of Monte Carlo integration in Bayesian inference. *Journal of Econometrics* 38: 73–90.

__________. 1991. Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints. *Computer Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface,* pp. 571–78. Alexandria, VA: American Statistical Association.

__________. 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Proceedings of the Fourth Valencia International Meeting on Bayesian Statistics,* ed. J. O. Berger, J. M. Bernardo, A. P. Dawid, and A. F. M. Smith, pp. 169–94. Oxford: Oxford University Press.

__________. 1993. Bayesian treatment of the independent student-t linear model. *Journal of Applied Econometrics* 8:S19–S40.

__________. 1994. Priors for macroeconomic time series and their application. *Econometric Theory*, forthcoming.

__________; Keane, M; and Runkle, D. 1994. Statistical inference for multinomial multiperiod probit models. Working Paper. University of Minnesota.

Hajivassiliou, V. 1992. Simulating normal orthant probabilities: The effects of vectorization. Paper presented at the NSF Conference on Large Scale Computation in Economics, October.

Hajivassiliou, V., and McFadden, D. 1990. The method of simulated scores for the estimation of LDV models with an application to external debt crises. Cowles Foundation Discussion Paper 967.

Hajivassiliou, V.; McFadden, D.; and Ruud, P. 1992. Simulation of multivariate normal orthant probabilities: Methods and programs. Cowles Foundation Discussion Paper 1021, Yale University.

Kahaner, D. K. 1991. A survey of existing multi-dimensional quadrature routines. In *Statistical Multiple Integrations Proceedings of the AMS-IMS-SIAM Joint Summer Research Conference.* Providence: American Mathematical Society.

Keane, M. 1990. Four essays in empirical macro and labor economics. Ph.D. dissertation. Brown University.

__________. 1992. A note on identification in the multinomial probit model. *Journal of Business and Economic Statistics* 10:2, 193–200.

__________. 1994. A computationally practical simulation estimator for panel data. *Econometrica* 62:1, 95-116.

__________, and Elrod, T. 1994. Modelling heterogeneity and state dependence in consumer choice behavior. Working Paper. University of Alberta.

Lee, L. F. 1992. Asymptotic bias in maximum simulated likelihood estimation of discrete choice models. Working Paper. University of Michigan.

Lerman, S., and Manski, C. 1981. On the use of simulated frequencies to approximate choice probabilities. In *Structural analysis of discrete data with econometric applications,* ed. C. Manski and D. McFadden. Cambridge: MIT Press.

McCulloch, R., and Rossi, P. E. 1994. An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*, forthcoming.

McFadden, D. 1984. Qualitative response models. In *Handbook of Econometrics,* ed. Z. Grilliches and M. Intriligator, Vol. 2, pp. 1395–458. Amsterdam: North-Holland.

__________. 1989. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica* 57:5, 995–1026.

McFadden, D., and Ruud P. 1992. Estimation by simulation. Manuscript. University of California at Berkeley.

Pakes, A., and Pollard, D. 1989. Simulation and the asymptotics of optimization estimators. *Econometrica* 57:5, 1027–58.

Schervish, M., and Carlin, B. 1992. On the convergence of successive substitution sampling. *Journal of Computational and Graphical Statistics,* forthcoming.

Tanner, M. A., and Wong, W. H. 1987. The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association* 82: 528–50.

Tierney, L. 1991. Markov chains for exploring posterior distributions. University of Minnesota School of Statistics Technical Report 560.

Zellner, A. 1962. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association* 57: 348–68.

___________. 1971. *An introduction to Bayesian inference in econometrics.* New York: Wiley.

___________, and Min, C. 1992. Gibbs sampler convergence criteria (GSC^2). Manuscript. Graduate School of Business, University of Chicago.