ELSEVIER

Journal of Econometrics 80 (1997) 125–165

JOURNAL OF Econometrics

Statistical inference in the multinomial multiperiod probit model

John F. Geweke[a,*], Michael P. Keane[a], David E. Runkle[b]

[a] *Department of Economics, University of Minnesota, Minneapolis, MN 55455, USA*
[b] *Federal Reserve Bank of Minneapolis, Research Department, 250 Marquette Avenue, Minneapolis, Minnesota 55401-0291, USA*

(Received 1 October 1994; received in revised form July 1996)

Abstract

Statistical inference in multinomial multiperiod probit models has been hindered in the past by the high dimensional numerical integrations necessary to form the likelihood functions, posterior distributions, or moment conditions in these models. We describe three alternative estimators, implemented using simulation-based approaches to inference, that circumvent the integration problem: posterior means computed using Gibbs sampling and data augmentation (GIBBS), simulated maximum likelihood (SML) estimation using the GHK probability simulator, and method of simulated moment (MSM) estimation using GHK. We perform a set of Monte-Carlo experiments to compare the sampling distributions of these estimators. Although all three estimators perform reasonably well, some important differences emerge. Our most important finding is that, holding simulation size fixed, the relative and absolute performance of the classical methods, especially SML, gets worse when serial correlation in disturbances is strong. In data sets with an AR(1) parameter of 0.50, the RMSEs for SML and MSM based on GHK with 20 draws exceed those of GIBBS by 9% and 0%, respectively. But when the AR(1) parameter is 0.80, the RMSEs for SML and MSM based on 20 draws exceed those of GIBBS by 79% and 37%, respectively, and the number of draws needed to reduce the RMSEs to within 10% of GIBBS are 160 and 80 respectively. Also, the SML estimates of serial correlation parameters exhibit significant downward bias. Thus, while conventional wisdom suggests that 20 draws of GHK is 'enough' to render the bias

* Corresponding author.

We wish to thank Peter Rossi for useful discussions and Donna Boswell for research assistance, but we are entirely responsible for any errors. Geweke's work was supported in part by National Science Foundation Grant SES-9210070. We thank the Federal Reserve Bank of Minneapolis for its support of this research. 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.

PII S0304-4076(97)00005-5

and noise induced by simulation negligible, our results suggest that much larger simulation sizes are needed when serial correlation in disturbances is strong. © 1997 Elsevier Science S.A.

Keywords: Bayesian inference; Discrete choice; Gibbs sampling; Method of simulated moments; Simulated maximum likelihood; Panel data
JEL classification: C35; C15

1. Introduction

Discrete economic choices are often made repeatedly over several time periods. Examples include the choice of which brand of a frequently purchased product category to buy on each successive purchase occasion and which of several industries or occupations to work in during each year of one's life. A multinomial multiperiod probit (MMP) model can be a reasonable framework for studying choice behavior in such situations. However, the very high dimensional integrations necessary to form the likelihood function, posterior distribution, or moment conditions for inference in the MMP model have until recently precluded its application. Rapid advances in simulation-based approaches to inference (McFadden, 1989; Pakes and Pollard, 1989; Keane, 1994a; McCulloch and Rossi, 1994) have now made both classical and Bayesian inference feasible. These advances have led to several interesting applications of the MMP model. These include sequential models of the decision to work (Keane, 1994a), brand choice (Elrod and Keane, 1995, Keane, 1994b, McCulloch and Rossi, 1994), choice of residential location (Hajivassiliou et al., 1996), and the probability a country will default on loans (Hajivassiliou and McFadden, 1994).

Despite this burgeoning list of applications, there has been no systematic comparison of the sampling distributions of alternative estimators in the MMP model in samples representative of these applications. The goal of the present paper is to provide such a comparison. First, we describe estimators based on three alternative approaches to inference: simulated maximum likelihood (SML) estimation using the Geweke–Hajivassiliou–Keane (GHK) recursive probability simulator, method of simulated moment (MSM) estimation using the GHK simulator, and posterior means computed using Gibbs sampling and data augmentation (GIBBS). We perform a set of Monte-Carlo experiments to compare the sampling distributions of the respective estimators. The experimental design allows the impact of three important features of the data on the performance of the methods to be assessed: (1) serial correlation of the random components of utility, (2) serial correlation of the exogenous variables, and (3) contemporaneous cross-alternative correlations of the random components of utility.

Besides features of the data, it is also important to consider how simulation size affects results. In the first set of experiments we hold the number of draws used to implement the GHK probability simulator fixed at 20. We choose 20 because conventional wisdom suggests this number is sufficient to render the bias intrinsic to the SML estimator negligible. For instance, Börsch-Supan and Hajivassiliou (1993) conclude that 'In our Monte Carlo experiment, 20 replications were sufficient to produce a negligible bias', and this conclusion has been influential. In fact, existing applications of SML generally use 20 or fewer draws. In a second set of experiments, we examine how performance of the classical estimators is affected by the number of draws used to implement the GHK probability simulator. In both experiments we set the number of cycles of the Gibbs sampler at 5000, since we find that this is sufficient to render the simulation noise in the posterior means very small as a fraction of root mean square error (RMSE).

Although all three estimators perform reasonably well in our experiments, some important differences emerge. Our most important finding is that, holding simulation size fixed, the relative and absolute performance of the classical methods, especially SML, gets worse when serial correlation in disturbances is strong. Consider the RMSEs of the SML and MSM point estimates and GIBBS posterior means around the data generating parameter values. In data sets with an AR(1) parameter of 0.50, the RMSEs for SML and MSM based on GHK with 20 draws exceed those of GIBBS by 9% and 0%, respectively. But when the AR(1) parameter is 0.80, the RMSEs for SML and MSM based on 20 draws exceed those of GIBBS by 79% and 37%, respectively, and the number of draws needed to reduce the RMSEs to within 10% of GIBBS are 160 and 80 respectively. Furthermore, the SML estimates of serial correlation parameters exhibit significant downward bias and this becomes insignificant only when 160 to 320 draws are used. Thus, contrary to conventional wisdom, 20 draws is not nearly 'enough' when serial correlation is strong.

This is the first systematic study of the performance of simulation-based approaches to inference in the MMP model in representative samples. There are five precursors of this work that are worth noting. Börsch-Supan and Hajivassiliou (1993) considered the distribution of the SML estimates of a single slope parameter in a cross-section trinomial probit model (with all other parameters held fixed at true values) using a single artificial data set, but varying the draws used in constructing the GHK simulator. McCulloch and Rossi (1994) provide Gibbs sampling data augmentation algorithms for the cross-section and panel probit models with random coefficients, but they do not allow for serial correlation in disturbances or compare sampling distributions of alternative estimators. Geweke et al. (1994) compared the sampling distributions of the SML, MSM and GIBBS estimators in the single-period multinomial probit model. Keane (1994a) studied the sampling distributions of the MSM and SML estimators based on the GHK probability simulator in the multiperiod probit model.

However, he only considered binomial probit models, did not consider Bayesian methods, and did not evaluate the influence of simulation size on the performance of the alternative methods. Hajivassiliou and McFadden (1994) study sampling distributions of SML and simulated score estimators in the multiperiod binomial probit model.

In Section 2 we describe the MMP model. Section 3 describes the SML and MSM estimators. In Section 4 we describe the GIBBS estimator. Section 5 lays out the design of our Monte-Carlo study, Section 6 presents results, and Section 7 concludes.

2. The model

Assume that agents choose among a set of J mutually exclusive alternatives in each of T time periods. If individual i chooses alternative j at time t, he/she derives utility

$$U_{ijt} = X'_{ijt}\beta_j + \varepsilon_{ijt} \quad (j = 1, \ldots, J;\ t = 1, \ldots, T),$$

where X_{ijt} is a $p \times 1$ vector of exogenous variables, β_j is a $p \times 1$ vector of corresponding coefficients, and ε_{ijt} is a random shock to utility that is known to the agent but not to the econometrician. Choice j is made at time t if $U_{ijt} > U_{ikt}$ for all $k \neq j$. The econometrician observes the choice

$$d_{ijt} = \begin{cases} 1 \text{ if } i \text{ chooses } j \text{ at time } t \\ 0 \text{ otherwise,} \end{cases}$$

but not the utility of any choice. The MMP model is obtained by assuming

$$\varepsilon_i \equiv (\varepsilon_{i11}, \ldots, \varepsilon_{iJ1}, \ldots, \varepsilon_{i1T}, \ldots, \varepsilon_{iJT})' \sim \text{IIDN}(0,\Sigma), \quad \Sigma = [\sigma_{jk}].$$

Since choices only depend on utility differences, it is conventional to measure utility *relative* to alternative J. Since the scale of utilities is indeterminate, it is also conventional to normalize by setting the variance of the error term corresponding to the first alternative in the transformed model equal to one. Thus, we define

$$\begin{aligned} U^*_{ijt} &= (U_{ijt} - U_{iJt})(\sigma_{11} + \sigma_{JJ} - 2\sigma_{1J})^{-1/2} \\ &= [(X'_{ijt}\beta_j - X'_{iJt}\beta_J) + (\varepsilon_{ijt} - \varepsilon_{iJt})](\sigma_{11} + \sigma_{JJ} - 2\sigma_{1J})^{-1/2} \\ &= X^{*\prime}_{ijt}\beta^*_j + \varepsilon^*_{ijt}\ (j = 1, \ldots, J;\ t = 1, \ldots, T), \end{aligned} \tag{2.1}$$

where $X^*_{ijt}\ (j = 1, \ldots, J)$ is the appropriate transformation of $X_{ijt}\ (j = 1, \ldots, J)$ and $\beta^*_j\ (j = 1, \ldots, J)$ is the appropriate transformation of $\beta_j (j = 1, \ldots, J)$. (Notice that $U^*_{iJt} = 0$ and $\varepsilon^*_{iJt} = 0$.) We further define

$$\begin{aligned} &\varepsilon^*_i = (\varepsilon^*_{i11}, \ldots, \varepsilon^*_{i,J-1,1}, \ldots, \varepsilon^*_{i1T}, \ldots, \varepsilon^*_{i,J-1,T})' \\ &\varepsilon^*_i \sim \text{IIDN}(0,\Sigma^*), \quad \Sigma^* = [\sigma^*_{jk}], \end{aligned} \tag{2.2}$$

where Σ^* is the corresponding appropriate transformation of Σ; by construction, $\sigma_{11}^* = 1$.

In the notation of the transformed model, choice j is made at time t if

$$U_{ijt}^* > U_{ikt}^* \text{ for all } k \neq j \ (j = 1, \ldots, J). \tag{2.3}$$

In order to have a compact notation for the sequence of choices observed for person i, define

$$d_{it} = (d_{i1t}, \ldots, d_{iJt}), \quad d_i = (d_{i1}, \ldots, d_{it}), \quad \text{and} \quad j_{it} = \{j | d_{ijt} = 1\}. \tag{2.4}$$

If $P(d_i)$ denotes the probability that i chooses the sequence d_i,

$$P(d_i) = P(U_{ij_{it}t}^* > U_{ikt}^* \ \forall k \neq j_{it}, \ t = 1, \ldots, T)$$

$$= P[\varepsilon_{ij_{it}t}^* - \varepsilon_{ikt}^* > X_{ikt}^{*\prime}\beta_k^* - X_{ij_{it}t}^{*\prime}\beta_{j_{it}}^* \ \forall k \neq j_{it}, \quad (t = 1, \ldots, T)].$$

If the ε_{ijt}^* are serially independent, then this is the product of T integrals each of dimension $J - 1$. However, if the ε_{ijt}^* are serially correlated, this is in general a $T(J-1)$ variate integral. As T and/or J grow, inference requiring exact evaluation of such integrals rapidly becomes infeasible. Much of the earlier work on the MMP model sought to avoid this problem by imposing low-order factor structures on Σ^*. For example, if a random effects structure is imposed, the order of integration is reduced to $2(J-1)$. The goal of simulation-based inference is to allow a richer covariance structure to be used.

In this paper we consider a special case of the model (2.1)–(2.4) in which the ε_{ijt}^* are stationary first-order autoregressive [AR(1)] processes and in which the X_{ijt}^* are divided into two sets of covariates: a set $\tilde{X}_{ijt}^*$ that is constant across alternatives (which can be thought of as containing characteristics of agent i) and a set Z_{ijt}^* that varies across alternatives (which can be thought of as containing attributes of alternatives, such as price or quality), but for which the corresponding coefficients are restricted to be equal across alternatives.

These decisions are motivated by a desire to study models that are practical. Note that even if the high-order integration problem can be solved by simulation techniques, unless J and T are both quite small it is not feasible to estimate an unrestricted Σ^* matrix which would contain $T^2(J-1)^2/2$ free parameters. This motivates our decision to study models in which the errors follow a stationary AR(1) process. Our partitioning of the covariates into two types is motivated not only by a desire to imitate applications, but also by the fact that likelihood surfaces in the multinomial probit model tend to be very flat unless one includes covariates that vary across alternatives (see Keane, 1990).

We next set out notation for the specific MMP model used in our experiments. Partition each coefficient vector $\beta_j^{*\prime} = (\tilde{\beta}_j^{*\prime}, \gamma')$ reflecting cross-equation constraints of the form employed in the experiments, and conformably partition

$X_{ijt}^{*\prime} = (\tilde{X}_{ijt}^{*\prime}, Z_{ijt}^{*\prime})$. Further define the matrices

$$\tilde{X}_{it}^{*} = \begin{bmatrix} \tilde{X}_{i1t}^{*\prime} & 0 & \cdots & 0 \\ 0 & \tilde{X}_{i2t}^{*\prime} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_{iLt}^{*\prime} \end{bmatrix}, \quad R = \begin{bmatrix} \rho_1 & 0 & \cdots & 0 \\ 0 & \rho_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \rho_L \end{bmatrix},$$

$$Z_{it}^{*} = \begin{bmatrix} Z_{i1t}^{*\prime} \\ Z_{i2t}^{*\prime} \\ \vdots \\ Z_{iLt}^{*\prime} \end{bmatrix},$$

where $L = J - 1$ and $|\rho_j| < 1$ $(j = 1, \ldots, L)$. Conformably define

$$\varepsilon_{it}^{*} = \begin{bmatrix} \varepsilon_{i1t} \\ \vdots \\ \varepsilon_{iLt} \end{bmatrix}, \quad U_{it}^{*} = \begin{bmatrix} U_{i1t}^{*} \\ \vdots \\ U_{iLt}^{*} \end{bmatrix}, \quad \tilde{\beta}^{*} = \begin{bmatrix} \tilde{\beta}_1^{*} \\ \vdots \\ \tilde{\beta}_L^{*} \end{bmatrix}, \quad \rho = \begin{bmatrix} \rho_1 \\ \vdots \\ \rho_L \end{bmatrix}.$$

In matrix notation, the model is then

$$U_{it}^{*} = \tilde{X}_{it}^{*}\tilde{\beta}^{*} + Z_{it}^{*}\gamma + \varepsilon_{it}^{*}.$$

The disturbances ε_{it}^{*} follow stationary AR(1) processes:

$$\varepsilon_{it}^{*} = R\varepsilon_{i,t-1}^{*} + v_{it}, \quad v_{it} \sim \text{IIDN}(0, \Psi), \quad \Psi = [\psi_{jk}].$$

Thus, the v_{it} are serially uncorrelated but correlated across alternatives. With this structure, $\sigma_{jk}^{*} = \psi_{jk}/(1 - \rho_k\rho_j)$. The assumption that R is diagonal is specific to the normalization on choice J in (2.1). In general, if R is diagonal for the given normalization, it will not be diagonal for alternative normalizations. The diagonality assumption made here will be most appealing when choice J is a baseline decision, such as a no purchase option in a brand choice model or a no work option in an occupational choice model, for which it is reasonable to assume utility is nonrandom.

3. Classical approaches to inference

3.1. Simulation of choice sequence probabilities

Classical estimators for the MMP model rely on Monte-Carlo simulation of the choice sequence probabilities $P(d_i)$ and substitution of these simulated probabilities into likelihood functions or moment conditions. In an extensive

study of alternative methods for simulation of multinomial orthant probabilities, Hajivassiliou et al. (1992) conclude that the GHK probability simulator, due to Keane (1990), Geweke (1991), and Hajivassiliou and McFadden (1994), is the most accurate of all methods considered. Geweke et al. (1994), in a Monte-Carlo study of alternative simulation estimators in the single-period multinomial probit model, concluded that classical methods based on GHK substantially outperformed classical methods based on kernel smoothed probability simulators. For these reasons, we rely exclusively on GHK to simulate choice probabilities when implementing classical estimators in this paper. In Appendix A we describe how to apply the GHK algorithm to simulation of choice sequence probabilities in the MMP model of Section 2. Below, we let $\hat{P}_{\mathrm{GHK}}(d_i|\beta^*, \Sigma^*, X^*)$ denote the GHK simulator of the probability of choice sequence d_i.

3.2. Classical estimation methods

The two classical estimation methods we consider are simulated maximum likelihood (SML) and method of simulated moments (MSM). The SML estimator maximizes the simulated log-likelihood function, which is obtained simply by substituting GHK simulators of choice sequence probabilities into the log-likelihood function:

$$L(\beta^*, \Sigma^*) = \sum_{i=1}^{N} \log \hat{P}_{\mathrm{GHK}}(d_i|\beta^*, \Sigma^*, X^*).$$

The SML estimator is consistent if $M/(N)^{1/2} \to \infty$ as $N \to \infty$. (For proofs, see Lee, 1992, 1995; and Gourieroux and Monfort, 1993.)

Direct application of McFadden's (1989) MSM estimator to the MMP model would involve indexing all possible choice sequences $s = 1, \ldots, J^T$ and defining choice indicators $d_{is} = 1$ if i chooses sequence s and 0 otherwise. Then form the MSM estimator by solving the moment conditions:

$$\sum_{i=1}^{N} \sum_{s=1}^{J^T} W_{is}[d_{is} - \hat{P}_{\mathrm{GHK}}(d_{is}|\beta^*_{\mathrm{MSM}}, \Sigma^*_{\mathrm{MSM}}, X^*)] = 0.$$

This MSM estimator is consistent for fixed M. This direct approach is not feasible because of the computational burden involved in simulating probabilities of J^T sequences and forming J^T weights.

Keane (1990) proposed the computationally feasible alternative of factoring the sequence probabilities into transition probabilities and forming the alternative estimator:

$$\sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{j=1}^{J} W_{ijt}[d_{ijt} - \hat{P}_{\mathrm{GHK}}(d_{ijt}|d_{i1}, \ldots, d_{i,t-1}, \beta^*_{\mathrm{MSM}}, \Sigma^*_{\mathrm{MSM}}, X^*)] = 0,$$

where the transition probabilities are simulated using ratios of GHK-simulated choice probabilities,

$$\hat{P}_{\mathrm{GHK}}(d_{ijt}|d_{i1}, \ldots, d_{i,t-1}) \equiv \frac{\hat{P}_{\mathrm{GHK}}(d_{i1}, \ldots, d_{i,t-1}, d_{ijt})}{\hat{P}_{\mathrm{GHK}}(d_{i1}, \ldots, d_{i,t-1})}.$$

Although this gives a biased simulator of the transition probability, an MSM estimator of this form is consistent if $M/(N)^{1/2} \to \infty$ as $N \to \infty$ (see Keane, 1994a). In addition, Keane (1994a) finds in a Monte-Carlo study that it has small sample properties superior to SML for the multiperiod binomial probit model, especially when serial correlation is strong.

4. 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 MMP model by McCulloch and Rossi (1994). Our approach is similar, but differs in four respects: Here, all priors are proper whereas McCulloch and Rossi use improper priors for β^*; we include autoregressive error components; stationarity is enforced through data augmentation of presample random utilities, rather than through explicit restrictions on Σ^* (see step 2 below); and the coefficients of covariates are fixed rather than random.

To provide a description of the Gibbs algorithm in generic notation, let $\tilde{Y}$, θ and Y denote vectors of latent utilities, model parameters, and observed choice data, respectively. Let $p(\theta, \tilde{Y}|Y)$ denote the joint posterior density function for θ and $\tilde{Y}$ conditional on Y. Suppose there is a partition of the parameter vector θ into B subvectors, $\theta' = (\theta'_{(1)}, \ldots, \theta'_{(B)})$, such that the conditional posterior densities $p(\theta_{(i)}|\theta_{(j)}, j \neq i, \tilde{Y}, Y)$ and $p(\tilde{Y}|\theta, Y)$ are of sufficiently simple form that it is practical to draw random subvectors $\tilde{\theta}_{(i)}$ and $\tilde{Y}$ from these conditional densities. The Gibbs algorithm starts with an initial value $(\theta^{(0)}, \tilde{Y}^{(0)})$ in the support of $p(\theta, \tilde{Y}|Y)$, and then draws in turn each of the subvectors $\tilde{Y}, \theta_{(1)}, \ldots, \theta_{(B)}$ from the appropriate conditional density. After each draw, the corresponding initial value subvector is replaced by the new subvector, until after a complete iteration an updated vector $(\theta^{(1)}, \tilde{Y}^{(1)})$ is obtained. After the mth iteration we obtain the draw $(\theta^{(m)}, \tilde{Y}^{(m)})$. As m grows larger the sample of $(\theta, \tilde{Y})$ draws converges in distribution to the joint posterior distribution. Posterior means for the elements θ are then approximated using arithmetic averages of the corresponding draws.

To describe the implementation of the Gibbs algorithm in the MMP model, some minor changes and extension in notation are necessary. Let the U^*_{ijt}, X^*_{ijt}, β^*_j, and ε^*_{ijt} continue to denote the latent utilities, covariates, coefficients, and disturbances of the transformed model, respectively, except that the

transformation (2.1) is replaced by

$$\begin{aligned} U^*_{ijt} &= (U_{ijt} - U_{iJt}) \\ &= [(X'_{ijt}\beta_j - X'_{iJt}\beta_J) + (\varepsilon_{ijt} - \varepsilon_{iJt})] \\ &= X^{*\prime}_{ijt}\beta^*_j + \varepsilon^*_{ijt} \quad (j = 1, J; \quad t = 1, T). \end{aligned}$$

The values of the β^*_j change accordingly, as does $\Psi = \mathrm{var}(\varepsilon^*_{it} - R\varepsilon^*_{i,t-1})$. The matrix R is unaffected.

Since the restriction $\sigma^*_{11} = 1$ has not been imposed, the parameters at this point are unidentified. In order to achieve identification, the following proper prior distributions were employed throughout the experiments:

$$\bar{\beta}^*_j \sim N(0, I_T); \quad \gamma \sim N(0, I_T); \quad \rho_j \sim TN(0.5, 0.25); \; \Psi^{-1} \sim W(10I_L, 10).$$

The proper prior distribution for Ψ centers the ψ_{jj}, which otherwise would be identified only up to a scale factor, about 1.0. This in turn induces a proper posterior distribution for the $\bar{\beta}^*_j$ and γ, even if the priors for these coefficients were flat and improper. (This technique was introduced by McCulloch and Rossi (1994).) The posterior distribution of these parameters induces a posterior distribution on $\Sigma^* = \mathrm{var}(\varepsilon^*_{it})$, with $\sigma^*_{jk} = \psi_{jk}/(1 - \rho_j\rho_k)$. Using the normalization set forth in Section 2, parameters of interest are $\bar{\beta}^*_j(\sigma^*_{11})^{-1/2}$ $(j = 1, \ldots, L)$; $\gamma(\sigma^*_{11})^{-1/2}$; ρ_j for $j = 1, \ldots, L$; and the elements of the upper triangular matrix A^*, where $A^{*\prime}A^* = (\sigma^*_{11})^{-1}\Psi$. To make drawings from the posterior distributions of these functions it is necessary only to transform the drawings of the $\bar{\beta}^*_j$, γ, ρ_j, and Ψ. In the experiments we will see that posterior standard deviations are very small relative to the priors, and on this basis it is reasonable to conjecture that results would be quite similar for other diffuse but proper priors.

We employ a six-block Gibbs sampling data augmentation algorithm in which the blocks are: (1) the latent utilities U^*_{ijt}; (2) the presample values of the errors, ε^*_{ij0}; (3) the ρ_j; (4) the matrix Ψ; (5) the vector $\bar{\beta}^* = (\bar{\beta}^{*\prime}_1, \ldots, \bar{\beta}^{*\prime}_L)'$; and (6) the γ. Although McCulloch and Rossi (1994) describe the structure of a Gibbs sampling data augmentation algorithm for a MMP model with random effects, the complications introduced by our addition of autoregressive error components are sufficiently great (including the addition of the new blocks 2 and 3 and changes necessary in other blocks) that we provide a detailed description of the algorithm for the present model in Appendix B.

In our experiments, the first 200 Gibbs iterations were discarded to allow 'burn in' from the initial drawing from the prior distribution. Inspection of these iterations showed that parameter values moved from well outside the concentration of the posterior distribution to its concentration in fewer than 100 iterations. Arithmetic averages of the parameters of interest over the next $m = 5000$ iterations were used to approximate posterior means. The standard errors of these Monte-Carlo approximations were assessed as described in Geweke

(1992). Typically, this standard error was less than 10% of the posterior standard deviation for the experiments undertaken.

The approximate posterior means obtained via the Gibbs sampling data augmentation algorithm constitute the GIBBS estimators that we compare with the SML and MSM estimators. Posterior means, given the priors assumed, constitute well-defined estimators with sampling-theoretic properties. Thus, out approach here is deliberately frequentist. A truly subjective Bayesian would have no reason to entertain either the MSM or SML methods, and would have little interest in posterior means given priors of convenience.

5. Experimental design

In our Monte-Carlo experiments, we consider a three alternative model ($J = 3$) with $T = 10$. We construct 20 artificial data sets of size $N = 500$ using the data generating process:

$$U^*_{i1t} = 0.5 + 1X^*_{it} + 1Z^*_{it1} + \varepsilon^*_{i1t}$$

$$U^*_{i2t} = -1.2 + 1X^*_{it} + 1Z^*_{it2} + \varepsilon^*_{i2t}$$

and of course, $U^*_{i3t} = 0.0$ by our normalization. The random shocks to utility evolve according to

$$\varepsilon^*_{i1t} = \rho_1 \varepsilon^*_{i1,t-1} + v_{i1t}$$

$$\varepsilon^*_{i2t} = \rho_2 \varepsilon^*_{i2,t-1} + v_{i2t}$$

$$\begin{bmatrix} v_{i1t} \\ v_{i2t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ a^*_{12} & \sqrt{1-(a^*_{12})^2} \end{bmatrix} \begin{bmatrix} \eta_{i1t} \\ \eta_{i2t} \end{bmatrix}$$

with $\eta_{it} \sim \text{IIDN}(0, I(1-\rho^2))$. In all the generated data, $\rho_1 = \rho_2 = \rho$. In the notation of Section 2,

$$\beta^*_1 = (0.5, 1, 1)', \quad \beta^*_2 = (-1.2, 1, 1)'$$

$$X^*_{i1t1} = X^*_{i2t1} \equiv X^*_{it}, \quad X^*_{i1t2} \equiv Z^*_{i1t}, \quad X^*_{i2t2} \equiv Z^*_{i2t}$$

$$\beta^*_{13} = \beta^*_{23} \equiv \gamma,$$

where X^*_{ijtl} refers to the lth element of the X^*_{ijt} vector. The regressors are constructed as follows:

$$X^*_{it} = \phi\mu_i + (1-\phi^2)^{1/2}\omega_{it} \ (t = 1, \ldots, T)$$

$$Z^*_{ijt} = \phi\psi_{ij} + (1-\phi^2)^{1/2}\xi_{ijt} \ (j = 1,2;\ t = 1, \ldots, T)$$

with

$$0 < \phi < 1, \quad \mu_i \sim \text{IIDN}(0, 1), \quad w_{it} \sim \text{IIDN}(0, 1), \quad \psi_{ij} \sim \text{IIDN}(0, 1),$$
$$\xi_{ijt} \sim \text{IIDN}(0, 1).$$

In our first experiment, we set the number of draws used to form the GHK simulator at 20, and consider 12 different data structures given by the $3 \times 2 \times 2$ full factorial design,

$$\rho_1 = \rho_2 = 0.50 \text{ or } 0.80; \; a_{12}^* = 0.50 \text{ or } 0.80; \; \phi^2 = 0 \text{ or } 0.50 \text{ or } 0.80.$$

These correspond to 'low' and 'high' serial correlation and cross correlation in the random elements of utility and 'no', 'low', and 'high' serial correlation in the exogenous variables, respectively. In the second experiment we vary the number of draws used to implement GHK, using two of the these data structures (chosen as described in Section 6.2).

6. Results

6.1. Experiment 1 — effect of data structure on performance of the estimators

The results of the Monte-Carlo experiments based on the 12 different data structures are reported in Tables 1–12. For GIBBS we report: (1) the mean of the posterior means across the 20 replications, $\bar{\tilde{\theta}}$; (2) the RMSE of the posterior means around the data generating values; and (3) the mean of the posterior standard deviation across the 20 replications, $\overline{\text{PSD}}$. For SML and MSM we report three statistics for each parameter in each model: (1) the mean of the point estimates across the 20 replications, $\bar{\hat{\theta}}$; (2) the root mean square error (RMSE) of the point estimates around the data generating values; and (3) the mean of the asymptotic standard errors across the 20 replications, $\overline{\text{ASE}}$. In the remainder of this section, we compare the performance of the different estimators in each experiment, focussing on RMSE as the criterion of performance. We also examine the $\overline{\text{ASE}}$ and $\overline{\text{PSD}}$, because, for purposes of inference, it is desirable that ASE or PSD be similar to RMSE.

In Table 1 we consider 20 artificial data sets generated from the data structure in which $\rho_1 = \rho_2 = 0.50$, $a_{12}^* = 0.50$, and $\phi^2 = 0$. This is the case of low serial and cross correlations of the disturbances combined with no serial correlation in the covariates. For GIBBS and MSM, the $\bar{\tilde{\theta}}$ are close to the data generating values for all 9 model parameters. SML, on the other hand, while producing estimates close to the data generating values for most model parameters, exhibits severe bias in estimating the AR(1) parameters. In particular, the mean SML point estimate of ρ_2 is 0.376, while the data generating value is 0.500. If we use the empirical RMSEs divided by $(20)^{1/2}$ to form t-tests for the estimated

Table 1
$\mathrm{Corr}(\eta_{1t}, \eta_{2t}) = 0.5$, $\rho_1 = \rho_2 = 0.5$, $\mathrm{Corr}(x_t, x_{t-1}) = 0.0$

		Bayesian inference			MSM-GHK			SML-GHK		
θ	DGP	$\bar{\theta}$	RMSE	$\overline{\mathrm{PSD}}$	$\bar{\theta}$	RMSE	$\overline{\mathrm{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\mathrm{ASE}}$
a_{12}^*	0.500	0.470	0.085	0.080	0.520	0.086	0.088	0.566	0.110	0.075
a_{22}^*	0.866	0.884	0.053	0.065	0.880	0.056	0.070	0.933	0.084	0.059
ρ_1^*	0.500	0.496	0.023	0.032	0.504	0.024	0.031	0.455	0.050	0.028
ρ_2^*	0.500	0.475	0.040	0.061	0.480	0.048	0.065	0.376	0.131	0.051
β_{11}^*	0.500	0.492	0.025	0.036	0.499	0.025	0.032	0.502	0.024	0.030
β_{21}^*	−1.200	−1.210	0.065	0.083	−1.204	0.074	0.078	−1.231	0.079	0.077
β_{12}^*	1.000	0.988	0.024	0.037	0.995	0.021	0.032	0.993	0.022	0.032
β_{22}^*	1.000	0.974	0.059	0.058	1.005	0.056	0.054	1.012	0.057	0.051
γ^*	1.000	0.995	0.024	0.037	0.993	0.028	0.033	0.995	0.025	0.032

Table 2
$\mathrm{Corr}(\eta_{1t}, \eta_{2t}) = 0.5$, $\rho_1 = \rho_2 = 0.5$, $\mathrm{Corr}(x_t, x_{t-1}) = 0.5$

		Bayesian inference			MSM-GHK			SML-GHK		
θ	DGP	$\bar{\theta}$	RMSE	$\overline{\mathrm{PSD}}$	$\bar{\theta}$	RMSE	$\overline{\mathrm{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\mathrm{ASE}}$
a_{12}^*	0.500	0.457	0.069	0.092	0.505	0.088	0.124	0.597	0.131	0.088
a_{22}^*	0.866	0.872	0.070	0.077	0.854	0.093	0.081	0.922	0.096	0.068
ρ_1^*	0.500	0.506	0.022	0.025	0.509	0.025	0.026	0.476	0.034	0.023
ρ_2^*	0.500	0.489	0.037	0.050	0.496	0.036	0.057	0.416	0.091	0.043
β_{11}^*	0.500	0.493	0.018	0.032	0.499	0.018	0.031	0.502	0.017	0.028
β_{21}^*	−1.200	−1.208	0.090	0.111	−1.180	0.115	0.106	−1.235	0.104	0.102
β_{12}^*	1.000	0.989	0.029	0.042	0.995	0.029	0.040	0.997	0.029	0.038
β_{22}^*	1.000	0.947	0.091	0.080	0.974	0.104	0.084	1.016	0.087	0.072
γ^*	1.000	0.987	0.034	0.043	0.981	0.038	0.040	0.987	0.034	0.039

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

Table 3
$\mathrm{Corr}(\eta_{1t}, \eta_{2t}) = 0.5$, $\rho_1 = \rho_2 = 0.5$, $\mathrm{Corr}(x_t, x_{t-1}) = 0.8$

		Bayesian inference			MSM-GHK			SML-GHK		
θ	DGP	$\bar{\theta}$	RMSE	$\overline{\mathrm{PSD}}$	$\bar{\theta}$	RMSE	$\overline{\mathrm{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\mathrm{ASE}}$
α^*_{12}	0.500	0.438	0.113	0.096	0.535	0.137	0.137	0.574	0.122	0.085
α^*_{22}	0.866	0.860	0.071	0.079	0.861	0.097	0.085	0.931	0.101	0.070
ρ^*_1	0.500	0.507	0.025	0.025	0.507	0.029	0.026	0.483	0.032	0.024
ρ^*_2	0.500	0.494	0.054	0.045	0.490	0.064	0.052	0.429	0.086	0.040
β^*_{11}	0.500	0.488	0.027	0.034	0.498	0.029	0.034	0.503	0.027	0.031
β^*_{21}	− 1.200	− 1.193	0.111	0.115	− 1.186	0.124	0.112	− 1.243	0.115	0.104
β^*_{12}	1.000	0.974	0.041	0.046	0.987	0.038	0.044	0.991	0.033	0.041
β^*_{22}	1.000	0.939	0.093	0.077	0.982	0.100	0.088	1.006	0.065	0.070
γ^*	1.000	0.986	0.041	0.047	0.986	0.039	0.044	0.996	0.036	0.042

Table 4
$\mathrm{Corr}(\eta_{1t}, \eta_{2t}) = 0.5$, $\rho_1 = \rho_2 = 0.8$, $\mathrm{Corr}(x_t, x_{t-1}) = 0.0$

		Bayesian inference			MSM-GHK			SML-GHK		
θ	DGP	$\bar{\theta}$	RMSE	$\overline{\mathrm{PSD}}$	$\bar{\theta}$	RMSE	$\overline{\mathrm{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\mathrm{ASE}}$
α^*_{12}	0.500	0.506	0.053	0.079	0.567	0.121	0.107	0.620	0.138	0.067
α^*_{22}	0.866	0.927	0.089	0.084	0.892	0.109	0.097	1.012	0.161	0.064
ρ^*_1	0.800	0.795	0.013	0.027	0.808	0.022	0.018	0.755	0.048	0.014
ρ^*_2	0.800	0.771	0.039	0.047	0.790	0.042	0.037	0.680	0.124	0.027
β^*_{11}	0.500	0.496	0.025	0.041	0.500	0.033	0.042	0.499	0.030	0.033
β^*_{21}	− 1.200	− 1.205	0.066	0.085	− 1.176	0.090	0.084	− 1.218	0.062	0.070
β^*_{12}	1.000	0.985	0.026	0.038	0.985	0.028	0.037	0.990	0.023	0.030
β^*_{22}	1.000	0.972	0.060	0.056	0.989	0.061	0.056	1.003	0.057	0.046
γ^*	1.000	0.991	0.030	0.039	0.982	0.038	0.036	0.991	0.025	0.030

Table 5
$\mathrm{Corr}(\eta_{1t}, \eta_{2t}) = 0.5$, $\rho_1 = \rho_2 = 0.8$, $\mathrm{Corr}(x_t, x_{t-1}) = 0.5$

		Bayesian inference			MSM-GHK			SML-GHK		
θ	DGP	$\bar{\theta}$	RMSE	$\overline{\mathrm{PSD}}$	$\bar{\theta}$	RMSE	$\overline{\mathrm{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\mathrm{ASE}}$
a_{12}^*	0.500	0.455	0.084	0.088	0.529	0.132	0.155	0.606	0.140	0.076
a_{22}^*	0.866	0.888	0.072	0.080	0.830	0.096	0.108	1.001	0.157	0.073
ρ_1^*	0.800	0.798	0.016	0.020	0.805	0.020	0.016	0.768	0.036	0.013
ρ_2^*	0.800	0.783	0.026	0.035	0.791	0.039	0.038	0.702	0.104	0.025
β_{11}^*	0.500	0.497	0.018	0.041	0.501	0.029	0.041	0.501	0.030	0.033
β_{21}^*	−1.200	−1.191	0.096	0.111	−1.121	0.140	0.116	−1.233	0.110	0.097
β_{12}^*	1.000	0.991	0.028	0.044	0.995	0.037	0.043	1.000	0.031	0.038
β_{22}^*	1.000	0.932	0.101	0.075	0.960	0.091	0.090	0.996	0.082	0.066
γ^*	1.000	0.980	0.040	0.045	0.966	0.054	0.044	0.987	0.038	0.037

Table 6
$\mathrm{Corr}(\eta_{1t}, \eta_{2t}) = 0.5$, $\rho_1 = \rho_2 = 0.8$, $\mathrm{Corr}(x_t, x_{t-1}) = 0.8$

		Bayesian inference			MSM-GHK			SML-GHK		
θ	DGP	$\bar{\theta}$	RMSE	$\overline{\mathrm{PSD}}$	$\bar{\theta}$	RMSE	$\overline{\mathrm{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\mathrm{ASE}}$
a_{12}^*	0.500	0.418	0.139	0.099	0.590	0.250	0.218	0.569	0.148	0.079
a_{22}^*	0.866	0.870	0.064	0.097	0.863	0.132	0.137	1.003	0.169	0.084
ρ_1^*	0.800	0.795	0.015	0.018	0.802	0.018	0.018	0.772	0.033	0.014
ρ_2^*	0.800	0.781	0.029	0.033	0.785	0.034	0.037	0.717	0.090	0.024
β_{11}^*	0.500	0.485	0.025	0.043	0.497	0.032	0.047	0.497	0.028	0.037
β_{21}^*	−1.200	−1.189	0.125	0.132	−1.162	0.164	0.152	−1.255	0.146	0.117
β_{12}^*	1.000	0.959	0.063	0.053	0.975	0.055	0.059	0.974	0.058	0.045
β_{22}^*	1.000	0.920	0.109	0.084	0.993	0.112	0.126	0.987	0.092	0.076
γ^*	1.000	0.973	0.059	0.054	0.962	0.076	0.060	0.988	0.050	0.046

Table 7
$\mathrm{Corr}(\eta_{1t}, \eta_{2t}) = 0.8$, $\rho_1 = \rho_2 = 0.5$, $\mathrm{Corr}(x_t, x_{t-1}) = 0.0$

		Bayesian inference			MSM-GHK			SML-GHK		
θ	DGP	$\bar{\theta}$	RMSE	$\overline{\mathrm{PSD}}$	$\bar{\theta}$	RMSE	$\overline{\mathrm{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\mathrm{ASE}}$
a^*_{12}	0.800	0.754	0.073	0.079	0.822	0.096	0.079	0.823	0.072	0.061
a^*_{22}	0.600	0.638	0.057	0.065	0.587	0.052	0.046	0.624	0.049	0.042
ρ^*_1	0.500	0.488	0.025	0.032	0.502	0.028	0.032	0.446	0.058	0.028
ρ^*_2	0.500	0.476	0.045	0.057	0.503	0.053	0.054	0.417	0.098	0.044
β^*_{11}	0.500	0.493	0.024	0.034	0.496	0.025	0.031	0.501	0.024	0.029
β^*_{21}	− 1.200	− 1.228	0.061	0.088	− 1.202	0.072	0.075	− 1.213	0.068	0.070
β^*_{12}	1.000	0.987	0.027	0.035	0.990	0.028	0.032	0.990	0.025	0.031
β^*_{22}	1.000	0.973	0.052	0.056	0.998	0.055	0.050	0.994	0.051	0.047
γ^*	1.000	1.003	0.025	0.038	0.990	0.027	0.033	0.994	0.024	0.032

Table 8
$\mathrm{Corr}(\eta_{1t}, \eta_{2t}) = 0.8$, $\rho_1 = \rho_2 = 0.5$, $\mathrm{Corr}(x_t, x_{t-1}) = 0.5$

		Bayesian inference			MSM-GHK			SML-GHK		
θ	DGP	$\bar{\theta}$	RMSE	$\overline{\mathrm{PSD}}$	$\bar{\theta}$	RMSE	$\overline{\mathrm{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\mathrm{ASE}}$
a^*_{12}	0.800	0.706	0.116	0.096	0.766	0.098	0.117	0.812	0.095	0.075
a^*_{22}	0.600	0.612	0.045	0.062	0.564	0.061	0.054	0.603	0.046	0.048
ρ^*_1	0.500	0.500	0.022	0.025	0.508	0.027	0.026	0.469	0.040	0.023
ρ^*_2	0.500	0.494	0.028	0.049	0.517	0.040	0.052	0.448	0.064	0.040
β^*_{11}	0.500	0.494	0.019	0.031	0.497	0.019	0.029	0.501	0.020	0.027
β^*_{21}	− 1.200	− 1.189	0.074	0.105	− 1.145	0.112	0.103	− 1.183	0.082	0.090
β^*_{12}	1.000	0.988	0.030	0.042	0.990	0.029	0.039	0.993	0.030	0.037
β^*_{22}	1.000	0.932	0.094	0.077	0.958	0.090	0.081	0.973	0.071	0.066
γ^*	1.000	0.992	0.028	0.044	0.979	0.034	0.040	0.982	0.035	0.038

Table 9
$\mathrm{Corr}(\eta_{1t}, \eta_{2t}) = 0.8$, $\rho_1 = \rho_2 = 0.5$, $\mathrm{Corr}(x_t, x_{t-1}) = 0.8$

		Bayesian inference			MSM-GHK			SML-GHK		
θ	DGP	$\bar{\theta}$	RMSE	$\overline{\mathrm{PSD}}$	$\bar{\theta}$	RMSE	$\overline{\mathrm{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\mathrm{ASE}}$
a^*_{12}	0.800	0.708	0.106	0.089	0.799	0.080	0.127	0.762	0.081	0.072
a^*_{22}	0.600	0.638	0.059	0.067	0.586	0.057	0.059	0.634	0.064	0.052
ρ^*_1	0.500	0.502	0.023	0.024	0.506	0.026	0.026	0.473	0.037	0.024
ρ^*_2	0.500	0.483	0.044	0.042	0.480	0.061	0.049	0.444	0.072	0.040
β^*_{11}	0.500	0.491	0.027	0.033	0.496	0.026	0.032	0.499	0.027	0.029
β^*_{21}	− 1.200	− 1.210	0.095	0.111	− 1.165	0.098	0.109	− 1.212	0.103	0.091
β^*_{12}	1.000	0.973	0.044	0.046	0.983	0.038	0.043	0.982	0.037	0.040
β^*_{22}	1.000	0.930	0.089	0.077	0.969	0.070	0.083	0.959	0.070	0.063
γ^*	1.000	1.001	0.041	0.048	0.985	0.043	0.045	1.001	0.042	0.041

Table 10
$\mathrm{Corr}(\eta_{1t}, \eta_{2t}) = 0.8$, $\rho_1 = \rho_2 = 0.8$, $\mathrm{Corr}(x_t, x_{t-1}) = 0.0$

		Bayesian inference			MSM-GHK			SML-GHK		
θ	DGP	$\bar{\theta}$	RMSE	$\overline{\mathrm{PSD}}$	$\bar{\theta}$	RMSE	$\overline{\mathrm{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\mathrm{ASE}}$
a^*_{12}	0.800	0.784	0.055	0.076	0.834	0.121	0.097	0.868	0.097	0.056
a^*_{22}	0.600	0.672	0.086	0.080	0.577	0.073	0.058	0.649	0.069	0.042
ρ^*_1	0.800	0.786	0.017	0.029	0.804	0.018	0.018	0.748	0.055	0.014
ρ^*_2	0.800	0.769	0.038	0.051	0.800	0.041	0.032	0.721	0.087	0.022
β^*_{11}	0.500	0.499	0.025	0.040	0.501	0.030	0.042	0.502	0.028	0.032
β^*_{21}	− 1.200	− 1.230	0.067	0.089	− 1.177	0.087	0.081	− 1.202	0.070	0.064
β^*_{12}	1.000	0.992	0.026	0.038	0.991	0.031	0.035	0.989	0.027	0.029
β^*_{22}	1.000	0.976	0.052	0.057	0.989	0.061	0.053	0.991	0.048	0.042
γ^*	1.000	1.007	0.030	0.039	0.989	0.036	0.036	0.991	0.030	0.030

Table 11
$\mathrm{Corr}(\eta_{1t}, \eta_{2t}) = 0.8$, $\rho_1 = \rho_2 = 0.8$, $\mathrm{Corr}(x_t, x_{t-1}) = 0.5$

		Bayesian inference			MSM-GHK			SML-GHK		
θ	DGP	$\bar{\theta}$	RMSE	$\overline{\mathrm{PSD}}$	$\bar{\theta}$	RMSE	$\overline{\mathrm{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\mathrm{ASE}}$
α_{12}^*	0.800	0.705	0.109	0.088	0.802	0.135	0.143	0.826	0.101	0.067
α_{22}^*	0.600	0.672	0.086	0.081	0.571	0.087	0.070	0.661	0.087	0.053
ρ_1^*	0.800	0.793	0.016	0.020	0.803	0.018	0.016	0.763	0.040	0.013
ρ_2^*	0.800	0.774	0.031	0.040	0.801	0.030	0.034	0.726	0.080	0.024
β_{11}^*	0.500	0.499	0.019	0.039	0.501	0.032	0.040	0.501	0.025	0.032
β_{21}^*	− 1.200	− 1.225	0.086	0.104	− 1.147	0.130	0.111	− 1.206	0.097	0.088
β_{12}^*	1.000	0.993	0.031	0.043	0.991	0.035	0.043	0.994	0.036	0.037
β_{22}^*	1.000	0.926	0.094	0.075	0.977	0.087	0.083	0.966	0.066	0.062
γ^*	1.000	1.000	0.034	0.046	0.970	0.048	0.044	0.992	0.039	0.037

Table 12
$\mathrm{Corr}(\eta_{1t}, \eta_{2t}) = 0.8$, $\rho_1 = \rho_2 = 0.8$, $\mathrm{Corr}(x_t, x_{t-1}) = 0.8$

		Bayesian inference			MSM-GHK			SML-GHK		
θ	DGP	$\bar{\theta}$	RMSE	$\overline{\mathrm{PSD}}$	$\bar{\theta}$	RMSE	$\overline{\mathrm{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\mathrm{ASE}}$
α_{12}^*	0.800	0.669	0.143	0.093	0.860	0.188	0.204	0.770	0.108	0.070
α_{22}^*	0.600	0.701	0.121	0.106	0.585	0.091	0.088	0.705	0.131	0.063
ρ_1^*	0.800	0.791	0.017	0.018	0.801	0.018	0.018	0.767	0.037	0.014
ρ_2^*	0.800	0.773	0.033	0.034	0.788	0.030	0.035	0.736	0.070	0.022
β_{11}^*	0.500	0.488	0.025	0.042	0.500	0.031	0.045	0.499	0.028	0.036
β_{21}^*	− 1.200	− 1.256	0.126	0.146	− 1.157	0.149	0.150	− 1.260	0.151	0.106
β_{12}^*	1.000	0.965	0.062	0.054	0.983	0.053	0.057	0.977	0.054	0.045
β_{22}^*	1.000	0.909	0.114	0.087	0.984	0.101	0.113	0.955	0.091	0.069
γ^*	1.000	1.002	0.045	0.056	0.967	0.062	0.061	1.005	0.049	0.045

deviations of mean point estimates (or mean posterior means) from data generating values, highly significant biases are found for the SML estimates of all covariance matrix parameters (a_{12}^*, a_{22}^*, ρ_1, ρ_2). No significant biases are found for the MSM estimates. For the GIBBS estimates, marginally significant biases are found only for ρ_2 and β_{12}^*.

In a comparison of RMSEs, GIBBS has an edge over the classical methods. It produces the smallest RMSE for six of the nine model parameters. Exceptions are β_{12}^* and β_{22}^* for which the RMSEs of the MSM point estimates are smallest, and β_{11}^* for which the RMSE of the SML point estimates is smallest. Another clear pattern is that for MSM the RMSE and the ASE are in close agreement for most model parameters. But for SML, the ASE are substantially below the RMSE for the covariance matrix parameters. Interestingly, the ASE for MSM and the PSD for GIBBS are in very close agreement. Given that the RMSEs for GIBBS are generally lower than for MSM, this also means that for GIBBS the PSDs are generally a bit above the corresponding RMSEs.

Rather than describing Tables 2–12 with the same level of detail devoted to Table 1, we instead point out certain broad patterns. As we move across Tables 1–3, the serial correlation in the covariates is increasing (ϕ^2 increases from 0 to 0.50 to 0.80) while other things are held constant. For most model parameters, the RMSEs for all three methods have a tendency to rise as ϕ^2 increases. The exception involves the ρ, for which the RMSEs fall as ϕ^2 increases. It also appears that the RMSEs for the SML estimates improve relative to those for other methods as ϕ^2 increases.

In Tables 4–6 the AR(1) parameters are increased (ρ_1 and ρ_2 are set at 0.80). Again, as we move across Tables 4–6, serial correlation in the covariates is increasing. Comparing Tables 4–6 with Tables 1–3, we see that the increase in the ρ generally causes RMSEs to rise. This is especially true for MSM. But for MSM and GIBBS, the increase in the ρ causes the RMSEs for the ρ to fall. This is not true for SML. Again, as in Tables 1–3, the RMSEs rise as ϕ^2 increases.

In Tables 7–9 the degree of serial correlation in the disturbances is returned to the Table 1–3 level (with ρ_1 and ρ_2 being set at 0.50), but the cross correlation of the errors is increased (a_{12}^* is set at 0.80). There is no obvious impact on the overall level of the RMSEs as compared to Tables 1–3. However, there is a substantial relative improvement for SML in the $\phi^2 = 0$ case of Table 7, where it produces the best RMSE for seven of nine parameters. And there is a substantial relative improvement for MSM in the $\phi^2 = 0.80$ case of Table 9, where it produces the best RMSE for four of nine parameters. These improvements in relative performance for SML and MSM do not extend to other cases.

In Tables 10–12 the degrees of serial and cross correlation in the errors are both set at the 'high' level ($\rho_1 = \rho_2 = 0.80$, $a_{12}^* = 0.80$). A comparison of Tables 10–12 with Tables 4–6 isolates the impact of increasing cross correlation when serial correlation in the errors is fixed at the high level. This leads to a clear reduction in the RMSE for the ρ as estimated by SML, but not for other

methods. Comparison of Tables 10–12 with Tables 7–9 isolates the impact of increasing serial correlation in the errors when cross correlation is fixed at the high level. This causes an increase of the RMSE for all parameters and for all methods, except for the ρ, for which the RMSEs decrease for GIBBS and MSM, but not for SML.

In Table 13 we present a regression that summarizes the relative performance of the estimators. The dependent variable is the log RMSE of the estimates for a parameter in one of the experiments. The right hand side variables are intercept and dummies for parameter, method of inference, and different levels of the treatments (that is, the degrees of serial and cross correlation in the errors, and the degree of serial correlation in the regressors), along with interactions of the treatment dummies with an indicator for whether the parameter is a ρ and interactions of method of inference with parameter and treatment levels. The intercept in the regression corresponds to the base case of GIBBS estimates for γ in the model with $\rho = 0.50$, $a_{12}^{*} = 0.50$, and $\phi = 0$.

The coefficient on MSM-GHK of 0.165 indicates that the RMSEs for the MSM estimates of γ tend to be roughly 16% greater than the RMSEs of the GIBBS posterior means. For SML the corresponding estimate is roughly 4%. By looking at the interactions of the parameter dummies with the method of inference, we can determine if the relative performance of the methods in terms of RMSEs differs systematically across parameters. The parameter with SML-GHK interactions produces some striking results. The interactions involving ρ_1 and ρ_2 have coefficients of 0.767 and 0.909 and are significant at the 1% level. Thus, the RMSEs of the SML point estimates for the serial correlation parameters are roughly 81% and 95% greater than those of the GIBBS posterior means. Also significant are the interactions involving the cross-correlation parameters a_{12}^{*} and a_{22}^{*}, which are 0.191 and 0.280. Thus, the performance of SML relative to other methods deteriorates substantially for these parameters as well.

Also of interest are the coefficients on the treatment dummies. These were entered both individually and in interaction with a dummy (DEP = RHO) for whether the parameter is ρ_1 or ρ_2. This is because, as the above discussion of Tables 1–12 made clear, there are obvious differences in how the treatments affect the RMSEs for the ρ's vs. all other model parameters. Note that the estimated main effect for $\rho_1 = \rho_2 = 0.80$ is 0.122. This indicates that raising the AR(1) parameters from 0.50 to 0.80 causes the RMSEs for GIBBS to rise by roughly 12% for parameters other than the ρ. However, the interaction of the $\rho_1 = \rho_2 = 0.80$ dummy with the ρ parameter dummy has a coefficient estimate of -0.421. This indicates that for the ρ parameters, raising the serial correlation in the errors causes the RMSE for GIBBS to fall by roughly 30%.

An important result is that the interactions of $\rho_1 = \rho_2 = 0.80$ with MSM and SML are both significantly positive, at 0.129 and 0.142, respectively. This indicates that for MSM and SML the increases in RMSEs for model parameters

other than the ρ parameters when serial correlation is strong are about 25–26% (vs. 12% for GIBBS), while the drops in RMSEs for the ρ parameters are only about 13–14% (vs. 30% for GIBBS). Thus, the performance of MSM and SML relative to GIBBS deteriorates substantially as serial correlation is increased.

The main effects of the $\phi^2 = 0.50$ and $\phi^2 = 0.80$ treatments are 0.185 and 0.468, indicating that increasing serial correlation the regressors causes the RMSEs for GIBBS to rise. The only one of the interactions of the $\phi^2 = 0.50$ and $\phi^2 = 0.80$ treatments with MSM-GHK and SML-GHK that is significant is the interaction of $\phi^2 = 0.80$ with SML-GHK. This has a coefficient of -0.143, which indicates that the performance of SML relative to GIBBS improves as serial correlation in the regressors increases. The interaction of $\phi^2 = 0.80$ with MSM-GHK also has a negative coefficient of -0.057, but this is not significant.

6.2. *Experiment 2 — effect of simulation size*

In this experiment we consider the effect of the number of draws used to construct the GHK simulator on the performance of MSM and SML. In particular, we want to determine whether the generally higher RMSEs for SML and MSM, relative to GIBBS, can be attributed to an insufficient number of draws in the GHK algorithm. Thus, we consider MSM and SML estimators implemented using GHK simulators based on 10, 20, 40, 80, 160, 320, 640 and 1280 draws.

It is impractical to repeat the analysis using all 12 data structures used in experiment 1 and all 8 alternative simulation sizes. Instead, we use the two data structures that generate the lowest and highest RMSEs for SML and MSM, relative to GIBBS. The last eight rows of Table 13 indicate that the best case for the classical estimators relative to GIBBS is data structure # 9, with low serial correlation in disturbances, high serial correlation in regressors, and high cross correlation in disturbances. Conversely, the worst case for SML and MSM is data structure # 4, in which these correlations are high, low, and low, respectively.

In Table 14 we report the results for SML, using the same 20 artificial data sets generated with data structure # 9 as were used in experiment 1. For each simulation size, we also report the average (across parameters) of RMSE relative to that for the GIBBS estimates based on $m = 5000$ iterations. We do this because standard methods for evaluating the accuracy of posterior moments (see Geweke, 1992) indicate that, with $m = 5000$, further Gibbs iterations should produce negligible changes in RMSE. Thus, the RMSEs for GIBBS provide a reasonable benchmark.

The Table 14 results are roughly consistent with the conventional wisdom reported in Börsch-Supan and Hajivassiliou (1993) that SML based on GHK performs well using only 20 draws. The mean across all 9 model parameters of

Table 13
Root mean square error comparison

Covariate	Predicted RMSE		
	Coeff.	Std Err	t-ratio
Intercept	−3.633	0.064	−56.493
MSM-GHK	0.165	0.090	1.846
SML-GHK	0.043	0.090	0.483
a_{12}^*	0.960	0.075	12.876
a_{22}^*	0.704	0.075	9.444
ρ_1^*	−0.111	0.089	−1.250
ρ_2^*	0.530	0.089	5.942
β_{11}^*	−0.421	0.075	−5.643
β_{21}^*	0.904	0.075	12.118
β_{12}^*	−0.024	0.075	−0.326
β_{22}^*	0.846	0.075	11.342
$\text{Corr}(\eta_{1t}, \eta_{2t}) = 0.80$	0.009	0.037	−0.237
$\rho_1 = \rho_2 = 0.80$	0.122	0.037	3.303
$\phi^2 = 0.50$	0.185	0.045	4.095
$\phi^2 = 0.80$	0.468	0.045	10.384
$\text{Corr}(\eta_{1t}, \eta_{2t}) = 0.80 \times \text{DEP} = \text{RHO}$	0.051	0.049	1.043
$\rho_1 = \rho_2 = 0.80 \times \text{DEP} = \text{RHO}$	−0.421	0.049	−8.615
$\phi^2 = 0.50 \times \text{DEP} = \text{RHO}$	−0.383	0.060	−6.409
$\phi^2 = 0.80 \times \text{DEP} = \text{RHO}$	−0.530	0.060	−8.864
$a_{12}^* \times$ MSM-GHK	0.103	0.105	0.980
$a_{22}^* \times$ MSM-GHK	−0.047	0.105	−0.447
$\rho_1^* \times$ MSM-GHK	−0.021	0.105	−0.196
$\rho_2^* \times$ MSM-GHK	−0.034	0.105	−0.326
$\beta_{11}^* \times$ MSM-GHK	−0.015	0.105	−0.140
$\beta_{21}^* \times$ MSM-GHK	0.061	0.105	0.574
$\beta_{12}^* \times$ MSM-GHK	−0.181	0.105	−1.713
$\beta_{22}^* \times$ MSM-GHK	−0.195	0.105	−1.852
$a_{12}^* \times$ SML-GHK	0.191	0.105	1.812
$a_{22}^* \times$ SML-GHK	0.280	0.105	2.650
$\rho_1^* \times$ SML-GHK	0.767	0.105	7.274
$\rho_2^* \times$ SML-GHK	0.909	0.105	8.621
$\beta_{11}^* \times$ SML-GHK	0.110	0.105	1.043
$\beta_{21}^* \times$ SML-GHK	0.108	0.105	1.028
$\beta_{12}^* \times$ SML-GHK	−0.046	0.105	−0.434
$\beta_{22}^* \times$ SML-GHK	−0.167	0.105	−1.585
$\text{Corr}(\eta_{1t}, \eta_{2t}) = 0.80 \times$ MSM-GHK	−0.082	0.050	−1.652
$\rho_1 = \rho_2 = 0.80 \times$ MSM-GHK	0.129	0.050	2.597
$\phi^2 = 0.50 \times$ MSM-GHK	0.033	0.061	0.543
$\phi^2 = 0.80 \times$ MSM-GHK	−0.057	0.061	−0.938
$\text{Corr}(\eta_{1t}, \eta_{2t}) = 0.80 \times$ SML-GHK	−0.129	0.050	−2.585
$\rho_1 = \rho_2 = 0.80 \times$ SML-GHK	0.142	0.050	2.865
$\phi^2 = 0.50 \times$ SML-GHK	−0.018	0.061	−0.292
$\phi^2 = 0.80 \times$ SML-GHK	−0.148	0.061	−2.431

Note: Dependent variable is log(RMSE). DEP = RHO means that the dependent variable for the observation is the log(RMSE) for a parameter ρ_1 or ρ_2. $\varepsilon_{1,t} = \rho_1 \varepsilon_{1,t-1} + \eta_1$, $\varepsilon_{2,t} = \rho_2 \varepsilon_{2,t-1} + \eta_2$, $\varepsilon_{3,t} = 0$.

Table 14
Performance of SML for alternative simulation sizes, Data Set # 9

		10 Draws			20 Draws			40 Draws		
θ	DGP	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$
a^*_{12}	0.800	0.752	0.094	0.066	0.762	0.081	0.072	0.788	0.072	0.078
a^*_{22}	0.600	0.673	0.087	0.052	0.634	0.064	0.052	0.616	0.055	0.051
ρ_1	0.500	0.456	0.052	0.023	0.473	0.037	0.024	0.487	0.027	0.024
ρ_2	0.500	0.411	0.101	0.038	0.444	0.072	0.040	0.456	0.061	0.041
β^*_{11}	0.500	0.500	0.025	0.029	0.499	0.027	0.030	0.500	0.026	0.030
β^*_{21}	− 1.200	− 1.254	0.116	0.090	− 1.212	0.103	0.091	− 1.190	0.099	0.094
β^*_{12}	1.000	0.983	0.036	0.038	0.982	0.037	0.039	0.981	0.037	0.040
β^*_{22}	1.000	0.965	0.073	0.061	0.959	0.070	0.063	0.963	0.067	0.065
ν	1.000	1.012	0.045	0.039	1.001	0.042	0.041	0.992	0.039	0.042
Mean ratio RMSE to Gibbs			1.310			1.093			0.971	

		80 Draws			160 Draws			320 Draws		
θ	DGP	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$
a^*_{12}	0.800	0.779	0.073	0.083	0.790	0.069	0.086	0.770	0.084	0.086
a^*_{22}	0.600	0.599	0.052	0.052	0.591	0.054	0.053	0.582	0.054	0.053
ρ_1	0.500	0.496	0.024	0.024	0.499	0.025	0.024	0.500	0.024	0.024
ρ_2	0.500	0.476	0.049	0.041	0.482	0.050	0.042	0.485	0.048	0.043
β^*_{11}	0.500	0.495	0.027	0.031	0.496	0.027	0.031	0.494	0.026	0.031
β^*_{21}	− 1.200	− 1.179	0.104	0.096	− 1.172	0.103	0.097	− 1.161	0.101	0.097
β^*_{12}	1.000	0.979	0.038	0.041	0.981	0.038	0.041	0.981	0.037	0.041
β^*_{22}	1.000	0.961	0.065	0.069	0.965	0.069	0.070	0.958	0.067	0.072
ν	1.000	0.989	0.044	0.043	0.987	0.042	0.043	0.987	0.043	0.043
Mean ratio RMSE to Gibbs			0.949			0.944			0.949	

		640 Draws			1280 Draws		
θ	DGP	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$
a_{12}^*	0.800	0.779	0.066	0.089	0.782	0.072	0.089
a_{22}^*	0.600	0.583	0.056	0.053	0.582	0.056	0.053
ρ_1	0.500	0.503	0.025	0.024	0.502	0.025	0.024
ρ_2	0.500	0.487	0.046	0.043	0.486	0.043	0.043
β_{11}^*	0.500	0.494	0.027	0.031	0.494	0.027	0.031
β_{21}^*	−1.200	−1.161	0.106	0.098	−1.162	0.105	0.098
β_{12}^*	1.000	0.980	0.038	0.042	0.981	0.037	0.042
β_{22}^*	1.000	0.962	0.065	0.071	0.964	0.064	0.071
γ	1.000	0.985	0.043	0.044	0.985	0.043	0.044
Mean ratio RMSE to Gibbs			0.939			0.931	

Table 15
Performance of MSM for alternative simulation sizes. Data Set # 9

		10 Draws			20 Draws			40 Draws		
θ	DGP	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$
a_{12}^*	0.800	0.746	0.116	0.146	0.799	0.080	0.127	0.778	0.073	0.109
a_{22}^*	0.600	0.574	0.064	0.061	0.586	0.057	0.059	0.586	0.059	0.057
ρ_1	0.500	0.505	0.034	0.028	0.506	0.026	0.026	0.504	0.025	0.025
ρ_2	0.500	0.494	0.061	0.052	0.480	0.061	0.049	0.481	0.056	0.045
β_{11}^*	0.500	0.492	0.031	0.032	0.496	0.026	0.032	0.497	0.028	0.031
β_{21}^*	− 1.200	− 1.155	0.097	0.113	− 1.165	0.098	0.109	− 1.160	0.105	0.106
β_{12}^*	1.000	0.980	0.037	0.043	0.983	0.038	0.043	0.984	0.037	0.042
β_{22}^*	1.000	0.946	0.080	0.089	0.969	0.070	0.083	0.963	0.066	0.077
γ	1.000	0.989	0.044	0.046	0.985	0.043	0.045	0.986	0.043	0.044
Mean ratio RMSE to Gibbs			1.115			0.995			0.982	

		80 Draws			160 Draws			320 Draws		
θ	DGP	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$
a_{12}^*	0.800	0.795	0.080	0.099	0.786	0.078	0.094	0.785	0.063	0.091
a_{22}^*	0.600	0.585	0.062	0.054	0.581	0.055	0.054	0.585	0.055	0.054
ρ_1	0.500	0.507	0.026	0.025	0.505	0.025	0.024	0.504	0.024	0.024
ρ_2	0.500	0.489	0.051	0.043	0.489	0.047	0.042	0.488	0.046	0.041
β_{11}^*	0.500	0.496	0.028	0.031	0.495	0.027	0.031	0.495	0.027	0.031
β_{21}^*	− 1.200	− 1.169	0.108	0.102	− 1.165	0.101	0.101	− 1.168	0.106	0.100
β_{12}^*	1.000	0.983	0.037	0.042	0.983	0.037	0.042	0.983	0.037	0.042
β_{22}^*	1.000	0.971	0.065	0.074	0.969	0.063	0.073	0.967	0.064	0.072
γ	1.000	0.986	0.042	0.044	0.987	0.043	0.043	0.988	0.043	0.043
Mean ratio RMSE to Gibbs			0.984			0.942			0.924	

		640 Draws			1280 Draws		
θ	DGP	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$
a_{12}^*	0.800	0.784	0.069	0.089	0.784	0.067	0.089
a_{22}^*	0.600	0.584	0.056	0.053	0.585	0.055	0.053
ρ_1	0.500	0.504	0.024	0.024	0.504	0.025	0.024
ρ_2	0.500	0.487	0.045	0.041	0.487	0.046	0.041
β_{11}^*	0.500	0.495	0.027	0.031	0.495	0.027	0.031
β_{21}^*	− 1.200	− 1.166	0.107	0.100	− 1.167	0.105	0.100
β_{12}^*	1.000	0.983	0.036	0.042	0.983	0.036	0.042
β_{22}^*	1.000	0.967	0.064	0.071	0.967	0.064	0.071
γ	1.000	0.987	0.043	0.043	0.987	0.043	0.043
Mean ratio RMSE to Gibbs			0.931			0.931	

Table 16
Performance of SML for alternative simulation sizes. Data Set # 4

		10 Draws			20 Draws			40 Draws		
θ	DGP	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$
a_{12}^*	0.500	0.631	0.150	0.064	0.620	0.138	0.067	0.606	0.125	0.069
a_{22}^*	0.866	1.042	0.189	0.064	1.012	0.161	0.064	0.987	0.142	0.066
ρ_1	0.800	0.719	0.083	0.015	0.755	0.048	0.014	0.775	0.029	0.014
ρ_2	0.800	0.617	0.189	0.028	0.680	0.124	0.027	0.717	0.087	0.026
β_{11}^*	0.500	0.503	0.024	0.030	0.499	0.030	0.033	0.497	0.028	0.035
β_{21}^*	− 1.200	− 1.237	0.075	0.070	− 1.218	0.062	0.070	− 1.207	0.069	0.072
β_{12}^*	1.000	0.997	0.023	0.029	0.990	0.023	0.030	0.989	0.024	0.031
β_{22}^*	1.000	1.016	0.061	0.045	1.003	0.057	0.046	0.999	0.050	0.046
γ	1.000	1.000	0.029	0.029	0.991	0.025	0.030	0.990	0.028	0.031
Mean ratio RMSE to Gibbs			2.345			1.791			1.473	

		80 Draws			160 Draws			320 Draws		
θ	DGP	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$
a_{12}^*	0.500	0.579	0.097	0.069	0.547	0.076	0.068	0.550	0.084	0.071
a_{22}^*	0.866	0.947	0.112	0.066	0.914	0.089	0.067	0.905	0.083	0.068
ρ_1	0.800	0.787	0.021	0.014	0.795	0.016	0.014	0.798	0.015	0.014
ρ_2	0.800	0.752	0.055	0.025	0.773	0.042	0.025	0.778	0.035	0.025
β_{11}^*	0.500	0.498	0.024	0.037	0.497	0.028	0.038	0.497	0.024	0.038
β_{21}^*	− 1.200	− 1.120	0.071	0.073	− 1.194	0.065	0.074	− 1.188	0.071	0.074
β_{12}^*	1.000	0.990	0.025	0.031	0.989	0.024	0.032	0.988	0.025	0.032
β_{22}^*	1.000	0.993	0.056	0.047	0.989	0.051	0.047	0.987	0.046	0.047
γ	1.000	0.992	0.030	0.031	0.989	0.032	0.032	0.988	0.030	0.032
Mean ratio RMSE to Gibbs			1.222			1.073			1.052	

θ	DGP	640 Draws			1280 Draws		
		$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$
a_{12}^*	0.500	0.544	0.075	0.072	0.539	0.072	0.072
a_{22}^*	0.866	0.897	0.077	0.069	0.890	0.079	0.070
ρ_1	0.800	0.801	0.014	0.014	0.800	0.013	0.015
ρ_2	0.800	0.787	0.028	0.026	0.789	0.029	0.025
β_{11}^*	0.500	0.496	0.025	0.039	0.497	0.026	0.039
β_{21}^*	−1.200	−1.193	0.072	0.075	−1.189	0.072	0.075
β_{12}^*	1.000	0.988	0.024	0.032	0.989	0.023	0.032
β_{22}^*	1.000	0.988	0.056	0.047	0.987	0.056	0.047
τ	1.000	0.989	0.030	0.032	0.989	0.029	0.032
Mean ratio RMSE to Gibbs			1.000			0.994	

Table 17
Performance of MSM for alternative simulation sizes, Data Set # 4

		10 Draws			20 Draws			40 Draws		
θ	DGP	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$
a_{12}^*	0.500	0.542	0.121	0.128	0.567	0.121	0.107	0.522	0.088	0.095
a_{22}^*	0.866	0.886	0.101	0.108	0.892	0.109	0.097	0.886	0.106	0.089
ρ_1	0.800	0.815	0.026	0.020	0.808	0.022	0.018	0.807	0.016	0.017
ρ_2	0.800	0.795	0.046	0.041	0.790	0.042	0.037	0.793	0.039	0.033
β_{11}^*	0.500	0.496	0.042	0.044	0.500	0.033	0.042	0.505	0.032	0.041
β_{21}^*	−1.200	−1.176	0.093	0.089	−1.176	0.090	0.084	−1.182	0.082	0.081
β_{12}^*	1.000	0.979	0.032	0.037	0.985	0.028	0.037	0.986	0.025	0.034
β_{22}^*	1.000	0.977	0.066	0.061	0.989	0.061	0.056	0.983	0.060	0.053
γ	1.000	0.977	0.048	0.038	0.982	0.038	0.036	0.988	0.031	0.035
Mean ratio RMSE to Gibbs			1.516			1.368			1.177	

		80 Draws			160 Draws			320 Draws		
θ	DGP	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$
a_{12}^*	0.500	0.543	0.093	0.084	0.546	0.087	0.078	0.539	0.083	0.074
a_{22}^*	0.866	0.892	0.092	0.080	0.894	0.098	0.077	0.883	0.089	0.073
ρ_1	0.800	0.805	0.015	0.016	0.805	0.014	0.015	0.804	0.015	0.015
ρ_2	0.800	0.791	0.034	0.029	0.792	0.031	0.028	0.794	0.028	0.026
β_{11}^*	0.500	0.500	0.026	0.040	0.497	0.026	0.040	0.500	0.025	0.040
β_{21}^*	−1.200	−1.194	0.080	0.079	−1.195	0.082	0.078	−1.191	0.079	0.077
β_{12}^*	1.000	0.990	0.022	0.034	0.989	0.023	0.033	0.990	0.024	0.033
β_{22}^*	1.000	0.990	0.055	0.051	0.989	0.058	0.051	0.987	0.062	0.050
γ	1.000	0.991	0.029	0.034	0.990	0.029	0.034	0.991	0.029	0.033
Mean ratio RMSE to Gibbs			1.089			1.075			1.055	

		640 Draws			1280 Draws		
θ	DGP	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$	$\bar{\theta}$	RMSE	$\overline{\text{ASE}}$
a_{12}^*	0.500	0.540	0.080	0.072	0.541	0.079	0.071
a_{22}^*	0.866	0.887	0.086	0.072	0.890	0.084	0.071
ρ_1	0.800	0.803	0.014	0.015	0.802	0.014	0.015
ρ_2	0.800	0.793	0.028	0.025	0.791	0.028	0.025
β_{11}^*	0.500	0.500	0.026	0.039	0.500	0.025	0.039
β_{21}^*	− 1.200	− 1.194	0.077	0.077	− 1.192	0.074	0.077
β_{12}^*	1.000	0.990	0.023	0.033	0.990	0.023	0.033
β_{22}^*	1.000	0.990	0.058	0.049	0.990	0.058	0.049
γ	1.000	0.991	0.029	0.033	0.991	0.029	0.033
Mean ratio RMSE to Gibbs			1.033			1.017	

the RMSE for SML is only 9% greater than GIBBS, and agreement between RMSE and $\overline{\text{ASE}}$ is generally good. The exceptions are ρ_1 and ρ_2, for which downward bias is noticeable and RMSEs are well above $\overline{\text{ASE}}$. Noticeable improvements in RMSEs and in agreement between RMSEs and ASEs for the autoregressive parameters are achieved by increasing the number of draws to 40 or 80, but little is gained by going beyond that.

Note that for the slope parameters β^* and γ there is little change in RMSE in going from 20 (or even 10) up to 1280 draws. Thus, the conventional wisdom that '20 draws is enough' may well stem from the tendency to focus on such parameters to the exclusion of the covariance structure parameters (in fact, recall that Börsch-Supan and Hajivassiliou (1993) held the covariance structure parameters fixed at true values in their study). But clearly, if a MMP model is to be used to model behavior, then the covariance parameters are every bit as important as the slope parameters.

The MSM results for data structure # 9 are reported in Table 15. When only 20 draws are used MSM performs even better than SML. The mean across all nine model parameters of the RMSE is essentially identical to that for GIBBS, and the agreement between RMSE and $\overline{\text{ASE}}$ is generally good. In contrast to SML, no downward bias is apparent for ρ_1 and ρ_2, even when only 10 draws are used. This is consistent with findings in Keane (1994). Again, it appears that increasing the number of draws beyond 20 leads to only minor improvements in RMSEs.

In Table 16 we report the results for SML using data structure # 4 – the worst case for MSM and SML. These results sharply contradict the conventional wisdom that '20 draws is enough'. Using only 20 draws, the mean RMSE for SML is roughly 79% greater than for GIBBS. Biases for all the covariance parameters are severe, especially for ρ_1 and ρ_2. 160 draws are needed to reduce the RMSE for SML to less than 10 percent greater than GIBBS, and 640 draws are needed to reduce it to the same level as for GIBBS. But even here, for the slope parameters it appears that the improvements in RMSE that are achieved by going past 20 draws are minor.

The MSM results for data structure # 4 are reported in Table 17. The performance of MSM in this worst case is much superior to that of SML. Using 20 draws, the mean RMSE for MSM is roughly 37% greater than for GIBBS. This is much better than the 79% figure achieved by SML. Even more importantly, there is no evidence of bias, even when only 10 draws are used. This is again consistent with findings in Keane (1994).

7. Conclusion

We have compared the sampling properties of the GIBBS estimator obtained by using Gibbs sampling and data augmentation to compute posterior means,

the MSM estimator using the GHK probability simulator, and the SML estimator using the GHK probability simulator. In a controlled Monte Carlo experiment we find that all three estimators perform reasonably well in point estimation of parameters of the data generating process in a three alternative 10-period multinomial multiperiod probit model. However, the relative performance of the methods varies as we vary the treatment variables in our experimental design.

Four important patterns emerge in the response of RMSEs to the design variables. First, RMSEs for all estimation methods and model parameters rise as serial correlation in either the errors or in the covariates is increased, except for the AR(1) parameters, for which RMSEs fall. Second, RMSEs of the SML and MSM estimates rise substantially relative to those of the GIBBS estimates as serial correlation in the errors increases. Third, increasing serial correlation in the covariates leads to a small relative decrease in the RMSEs of the SML estimates. Fourth, increasing cross correlation of the errors leads to relative declines in the RMSEs of the SML and MSM estimators.

Of these patterns, we feel that the most important to a user of these methods is the fact that, holding draws fixed, both the relative and absolute performance of the classical methods, especially SML, becomes worse as serial correlation in disturbances increases. In data sets with an AR(1) parameter of 0.50, the RMSEs for SML and MSM based on 20 draws exceed those of GIBBS by 9% and 0%, respectively. But when the AR(1) parameter is 0.80, the RMSEs for SML and MSM based on 20 draws exceed those of GIBBS by 79% and 37%, respectively, and the number of draws needed to reduce the RMSEs for SML and MSM to within 10% of GIBBS are 160 and 80, respectively.

In existing applications it is common for MSM or SML to be implemented using GHK simulators based on only 10 or 20 draws, and conventional wisdom seems to suggest that 20 draws is 'enough'. But our results suggest that when serial correlation in the errors is strong, substantial RMSE reductions can be achieved by going to 80 or 160 draws. And in the case of SML, substantial bias reductions can be achieved as well. The main weakness of SML relative to the other methods is in estimation of serial correlation parameters. When 20 draws are used, the SML estimates of the AR(1) coefficients are clearly biased downward and have RMSEs roughly twice as great as for other methods. This problem would lead to larger RMSEs in out of sample forecasts obtained using SML as compared to the other methods. It appears that from 80 to 320 draws must be used in order to render these biases insignificant.

For purposes of statistical inference it is desirable that ASE or PSD be similar to RMSE. There is generally close agreement between RMSEs and ASEs for MSM, and somewhat less close but still acceptable agreement between RMSEs and PSDs for GIBBS. But, when only 20 draws are used, RMSEs of SML point estimates exceed ASEs by 38% on average. This divergence is greater when serial correlation in errors is strong, and it can be reduced by using more draws.

What accounts for the increases in RMSEs for MSM and SML relative to GIBBS as serial correlation in the errors is increased while holding simulation size fixed? It stems from the different roles of simulation in the estimators. In MSM and SML, a set of nonlinear equations is solved iteratively. Equation evaluation at *each* iteration is approximate, but consistent in the number of GHK draws. As serial correlation increases, more draws are needed to maintain a given level of accuracy of the GHK approximation. In GIBBS, simulation-consistent approximation of posterior means comes about because, over the whole set of iterations, the simulator draws a sample representative of the posterior distribution. With increased disturbance serial correlation, a greater number of iterations is required to achieve a representative sample because the Gibbs sampler navigates the posterior more slowly. Thus, performance of each method should deteriorate as serial correlation in disturbances is increased, but the relative rate of deterioration is an empirical question. In our experiments, 5000 iterations of the Gibbs sampler provides good approximations to the posterior for both high and low degrees of serial correlation. In contrast, 20 draws for the GHK simulator provides comparable approximation accuracy when serial correlation is low, but the number must be made much greater when serial correlation is high.

Since we argue that RMSEs of the MSM and SML estimators based on GHK with 20 draws and GIBBS posterior means based on 5000 Gibbs iterations are similar when serial correlation is relatively weak (i.e., AR(1) parameters in the 0.50 range), but that roughly 80 to 320 draws for MSM and SML are needed to maintain comparable performance when serial correlation is strong (i.e., AR(1) parameters in the 0.80 range), it is important to consider how computation time for the various methods is affected by simulation size.

Unfortunately, it is essentially impossible to give absolute time comparisons for these methods for several reasons. For example, since most of the computation time for GIBBS is used in drawing the latent utilities, time for GIBBS increases slowly as the number of model parameters increases. But for SML and MSM, computation time is roughly proportional to the number of parameters, since most of the time in these methods is spent calculating derivatives and the initial weighting matrix, respectively. Also, relative timings for the methods will depend on the number of iterations necessary to achieve convergence for the classical methods, and the number of Gibbs iterations needed to achieve a desired level of numerical accuracy. In particular, for MSM the calculation of the initial weighting matrix is much more time consuming that subsequent iterations, while for SML all iterations take roughly equal time. And the number of iterations needed for convergence is particular to the application at hand. Finally, if one wishes to check sensitivity of results to simulation size, MSM and SML require that one redo the entire estimation using more draws, while with GIBBS one can simply restart the algorithm from end values and add additional iterations. Thus, monitoring numerical accuracy is much easier with GIBBS.

Despite these problems, it is informative to give some timing information based on our Monte Carlo experiments. The typical time for our GIBBS runs on an IBM 3090 mainframe was 816 cpu s. Timings for the classical methods is much more variable than for GIBBS, because the number of iterations needed to achieve convergence varies greatly across Monte Carlo data sets. On average, the number of iterations needed to achieve convergence for the data sets considered here is about 10. For SML, a regression of computation time per iteration on number of draws produced the result 5.2 + 0.74*Draws. Thus, the predicted times for ten iteration runs using 20, 80 and 160 draws are 200, 644 and 1236 cpu s, respectively. For MSM, a regression of marginal computation time per iteration on number of draws produced the result 2.7 + 0.54*Draws, while a regression of initial weighting matrix computation time on number of draws produced the result − 2.0 + 3.44*Draws. Thus, the predicted times for ten iteration runs using 20, 80 and 160 draws are 202, 732 and 1439 cpu s, respectively. We interpret these results as indicating that computation times are in the same ballpark for all three methods, given models of the size considered here and simulation sizes of the type one would typically employ. Thus, we conclude that computation time should not be a decisive factor in choosing among the methods.

If we had found that computation times for GIBBS greatly exceeded those for MSM or SML, it would make sense for a Bayesian practitioner to consider these methods as a way to approximate posterior means. That this is not the case implies that a Bayesian should have no interest in these alternatives – provided that Gibbs sampling data augmentation software is at hand.

Appendix A: application of the GHK simulator to the MMP model

Here we provide a description of the GHK simulator as applied to the simulation of choice probabilities in the MMP model described in Section 2. For a proof of the unbiasedness of the GHK simulator in general, see Börsch-Supan and Hajivassiliou (1993). To describe the GHK simulator it is useful to define some additional notation. Let

$$\tilde{U}^j_{ikt} = U^*_{ikt} - U^*_{ijt} \quad (j = 1, \ldots, J;\ t = 1, \ldots, T), \quad \tilde{\varepsilon}^j_{ikt} = \varepsilon^*_{ikt} - \varepsilon^*_{ijt}.$$

(Notice that $\tilde{U}^j_{ijt} = 0$ and $\tilde{\varepsilon}^j_{ijt} = 0$.) Choice j is made at t if the $J-1$ constraints $\tilde{U}^j_{ikt} < 0$ for all $k \neq j$ are satisfied. Further let

$$\tilde{\varepsilon}_{it}(-j) = (\tilde{\varepsilon}^j_{i1t}, \ldots, \tilde{\varepsilon}^j_{i,j-1,t}, \tilde{\varepsilon}^j_{i,j+1,t}, \ldots, \tilde{\varepsilon}^j_{iJt})' \quad \text{and}$$

$$\tilde{\varepsilon}(d_i) = (\tilde{\varepsilon}_{i1}(-j_{i1}), \ldots, \tilde{\varepsilon}_{iT}(-j_{iT}))'$$

where d_i is the choice vector defined in (2.4). Thus $\tilde{\varepsilon}(d_i) \sim \text{IIDN}(0, \tilde{\Sigma}(d_i))$, where $\tilde{\Sigma}(d_i)$ is the appropriate transformation of Σ^*.

Let $\tilde{A}(d_i)$ be the unique lower triangular Cholesky decomposition $\tilde{\Sigma}(d_i) = \tilde{A}(d_i)\tilde{A}(d_i)'$. Then $\tilde{\varepsilon}(d_i) = \tilde{A}(d_i)\tilde{\eta}(d_i)$, where (suppressing the i subscript) $\tilde{\eta}_t(-j_t) = (\eta_{1t}, \ldots, \eta_{j_t-1,t}, \eta_{j_t+1,t}, \ldots, \eta_{Jt})'$, $\tilde{\eta}(d_i) = (\tilde{\eta}_1(-j_1), \ldots, \tilde{\eta}_T(-j_T))'$, and $\eta_{ijt} \sim \text{IDN}(0, 1)$ for all i, j, t.

Finally, define $\tilde{U}^j_{ikt}(\tilde{\eta}^l_{11}, \ldots, \tilde{\eta}^l_{J,t-1}, \tilde{\eta}^l_{1t}, \ldots, \tilde{\eta}^l_{pt})$ as the value of $\tilde{U}^j_{ikt}$ when the random variables $(\tilde{\eta}_{11}, \ldots, \tilde{\eta}_{J,t-1}, \tilde{\eta}_{1t}, \ldots, \tilde{\eta}_{pt})$ are fixed at the *draw* $(\tilde{\eta}^l_{11}, \ldots, \tilde{\eta}^l_{J,t-1}, \tilde{\eta}^l_{1t}, \ldots, \tilde{\eta}^l_{pt})$. Note that for $p = k$ this is a number, and for $p < k$ this is a random variable. Then, the GHK simulator for the probability of the choice sequence $(d_{i1}, \ldots, d_{it})$, or equivalently $(j_{i1}, \ldots, j_{it})$, is constructed as follows (suppressing the i subscript):

Period 1:
Step:

(1) Draw η^l_{11} s.t. $\tilde{U}^{j_1}_{11}(\eta^l_{11}) < 0$

$\vdots$

$(j_1 - 1)$ Draw $\eta^l_{j_1-1,1}$ s.t. $\tilde{U}^{j_1}_{j_1-1,1}(\eta^l_{11}, \ldots, \eta^l_{j_1-1,1}) < 0$

(j_1) Skip $\eta^l_{j_1,1}$

$(j_1 + 1)$ Draw $\eta^l_{j_1+1,1}$ s.t. $\tilde{U}^{j_1}_{j_1+1,1}(\eta^l_{11}, \ldots, \eta^l_{j_1-1,1}, \eta^l_{j_1+1,1}) < 0$

$\vdots$

(J) Draw η^l_{J1} s.t. $\tilde{U}^{j_1}_{J1}(\eta^l_{11}, \ldots, \eta^l_{j_1-1,1}, \eta^l_{j_1+1,1}, \ldots, \eta^l_{J1}) < 0$

Period t:
Step:

(1) Draw η^l_{1t} s.t. $\tilde{U}^{j_t}_{1t}(\tilde{\eta}^l_{11}, \ldots, \tilde{\eta}^l_{J,t-1}, \eta^l_{1t}) < 0$

$\vdots$

$(j_t - 1)$ Draw $\eta^l_{j_t-1,t}$ s.t. $\tilde{U}^{j_t}_{j_t-1,t}(\tilde{\eta}^l_{11}, \ldots, \tilde{\eta}^l_{J,t-1}, \tilde{\eta}^l_{1t}, \ldots, \eta^l_{j_t-1,t}) < 0$

(j_t) Skip $\eta^l_{j_t,t}$

$(j_t + 1)$ Draw $\eta^l_{j_t+1,t}$ s.t. $\tilde{U}^{j_t}_{j_t+1,t}(\tilde{\eta}^l_{11}, \ldots, \tilde{\eta}^l_{J,t-1}, \eta^l_{1t}, \ldots, \eta^l_{j_t-1,t}, \eta^l_{j_t+1,t}) < 0$

$\vdots$

(J) Draw η^l_{Jt} s.t. $\tilde{U}^{j_t}_{Jt}(\tilde{\eta}^l_{11}, \ldots, \tilde{\eta}^l_{J,t-1}, \eta^l_{1t}, \ldots, \eta^l_{j_t-1,t}, \eta^l_{j_t+1,t}, \ldots, \eta^l_{Jt}) < 0$

and finally, construct:

$$\hat{P}_{\text{GHK}}(d_1, \ldots, d_t | \beta^*, \Sigma^*, X^*)$$
$$= \frac{1}{M}\sum_{l=1}^{M} P(\tilde{U}^{j_1}_{11} < 0) \prod_{k=2}^{j_t-1} P[\tilde{U}^{j_t}_{k1}(\eta^l_{11}, \ldots, \eta^l_{k-1,1}) < 0]$$

$$\times P[\tilde{U}^{j_1}_{j_1+1,1}(\eta^l_{11}, \ldots, \eta^l_{j_1-1,1}) < 0]$$

$$\prod_{k=j_1+2}^{J} P[\tilde{U}^{j_1}_{k1}(\eta^l_{11}, \ldots, \eta^l_{j_1-1,1}, \eta^l_{j_1+1,1}, \ldots, \eta^l_{k-1,1}) < 0]$$

$$\cdot \ldots \cdot P(\tilde{U}^{j_t}_{1t}(\tilde{\eta}^l_{11}, \ldots, \tilde{\eta}^l_{J_t-1}) < 0)$$

$$\times \prod_{k=2}^{j_t-1} P[\tilde{U}^{j_t}_{kt}(\tilde{\eta}^l_{11}, \ldots, \tilde{\eta}^l_{J_t-1}, \eta^l_{1t}, \ldots, \eta^l_{k-1,t}) < 0]$$

$$P[\tilde{U}^{j_t}_{j_t+1,t}(\tilde{\eta}^l_{11}, \ldots, \tilde{\eta}^l_{J_t-1}, \tilde{\eta}^l_{1t}, \ldots, \tilde{\eta}^l_{j_t-1,t}) < 0]$$

$$\prod_{k=j_t+2}^{J} P[\tilde{U}^{j_t}_{kt}(\tilde{\eta}^l_{11}, \ldots, \tilde{\eta}^l_{J_t-1}, \eta^l_{1t}, \ldots, \eta^l_{j_t-1,t}, \eta^l_{j_t+1,t}, \ldots, \eta^l_{k-1,t}) < 0].$$

Appendix B: a Gibbs sampling data augmentation algorithm for the MMP model

In order to describe the Gibbs sampling data augmentation algorithm, it is first useful to define the proper prior distributions in generic notation:

$$\tilde{\beta}^*_j \sim N(\bar{\beta}^*_j, V_{\tilde{\beta}^*_j}) \quad (j = 1, \ldots, J-1)$$

$$\gamma \sim N(\gamma, V_\gamma)$$

$$\rho_j \sim TN(\rho_j, \sigma^2_{\rho_j}) \quad (j = 1, \ldots, J-1),$$

where TN denotes truncation of the univariate normal distribution to the unit interval (0, 1);

$$\Psi^{*-1} \sim W(S^{-1}; \nu),$$

where W denotes the Wishart distribution. These priors are assumed mutually independent.

It is also useful to adopt the notational convention $U^*_{i0} = \varepsilon^*_{i0}$, $\tilde{X}^*_{i0} = [0]$, $Z^*_{i0} = [0]$, and to define three equivalent expressions for the probability density kernel of the U^*_{it}, which are

$$|\Psi|^{-NT/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}(\varepsilon^*_{it} - R\varepsilon^*_{i,t-1})'\Psi^{-1}(\varepsilon^*_{it} - R\varepsilon^*_{i,t-1})\right\}, \tag{B.1}$$

where $\varepsilon^*_{it} = U^*_{it} - X^*_{it}\tilde{\beta}^* - Z^*_{it}\gamma$; or

$$|\Psi|^{-NT/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}(\dot{U}^*_{it} - \dot{X}^*_{it}\tilde{\beta}^* - \dot{Z}^*_{it}\gamma)'\right.$$

$$\left.\times \Psi^{-1}(\dot{U}^*_{it} - \dot{X}^*_{it}\tilde{\beta}^* - \dot{Z}^*_{it}\gamma)\right\}, \tag{B.2}$$

where

$$\dot{U}_{it}^{*} = U_{it}^{*} - RU_{i,t-1}^{*}, \quad \dot{X}_{it}^{*} = \tilde{X}_{it}^{*} - R\tilde{X}_{i,t-1}^{*}, \quad \dot{Z}_{it}^{*} = Z_{it}^{*} - RZ_{i,t-1}^{*};$$

and

$$|\Psi|^{-NT/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}[U_{it}^{*} - RU_{i,t-1}^{*} - (\mu_{it} - R\mu_{i,t-1})]\,\Psi^{-1}\right.$$

$$\left.\times[U_{it}^{*} - RU_{i,t-1}^{*} - (\mu_{it} - R\mu_{i,t-1})]\right\}, \tag{B.3}$$

where $\mu_{it} = X_{it}^{*}\tilde{\beta}^{*} + Z_{it}^{*}\gamma$, each multiplied by the kernel density of the unconditional distribution of ε_{i0}^{*},

$$|V_0(R,\Psi)|^{-N/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\varepsilon_{i0}^{*\prime}[V_0(R,\Psi)]^{-1}\varepsilon_{i0}^{*}\right\}, \tag{B.4}$$

where $[V_0(R,\Psi)]_{jk} = \psi_{jk}/(1-\rho_j\rho_k)$.

The kernel of the posterior density function for β^{*}, γ, ρ, Ψ, and the U_{it}^{*} is the product of (B.1) or (B.2) or (B.3); (B.4); the density kernels for the prior distributions; and

$$\prod_{i=1}^{N}\prod_{t=1}^{T}H(U_{it}^{*}, d_{iT}), \tag{B.5}$$

where H is an indicator function for consistency of orderings and signs of the $U_{ij_t}^{*}$ $(j = 1, \ldots, L)$ with the observed choice vectors d_{it}. Using notation defined in Section 2, this is

$$H(U_{it}^{*}, d_{it}) = 1 \text{ iff } U_{it}^{*} \in \{U_{it}^{*} | U_{i,j_{it},t}^{*} > U_{ikt}^{*} \forall k \neq j_{it}\}.$$

A six-step Gibbs sampling-data augmentation algorithm is employed to construct draws from the posterior distribution. Initial values for $\tilde{\beta}^{*}$, γ, ρ, and Ψ are drawn from the prior distributions, and initially $U_{it}^{*} = 0$ $(i = 1, \ldots, N;$ $t = 1, \ldots, T)$.

Step 1: Drawing $U_{it}^{*}(i = 1, \ldots, n;\ t = 1, \ldots, T)$. The kernel density of the conditional distribution of the U_{it}^{*} is the product of (B.3) and (B.5). The conditional distribution of $U_t^{*} = (U_{1t}^{*\prime}, \ldots, U_{nT}^{*\prime})'$ is truncated normal. The conditional normal distribution is given by the relations

$$U_{i1}^{*} - \mu_{i1} - R\varepsilon_{i0}^{*} = v_{i1},$$

$$U_{i2}^{*} - \mu_{i2} - R(U_{i1}^{*} - \mu_{i1}) = v_{i2}$$

$$\vdots$$

$$U_{iT}^{*} - \mu_{iT} - R(U_{i,T-1}^{*} - \mu_{i,T-1}) = v_{iT}$$

from (B.3), which imply

$$\begin{bmatrix} I_T & 0 & 0 & \cdots & 0 & 0 & 0 \\ -R & I_T & 0 & \cdots & 0 & 0 & 0 \\ 0 & -R & I_T & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & & & \\ 0 & 0 & 0 & \cdots & -R & I_T & 0 \\ 0 & 0 & 0 & \cdots & 0 & -R & I_T \end{bmatrix} \begin{bmatrix} U_{i1}^* - \mu_{i1} \\ U_{i2}^* - \mu_{i2} \\ U_{i3}^* - \mu_{i3} \\ \vdots \\ U_{i,T-1} - \mu_{i,T-1} \\ U_{iT}^* - \mu_{iT} \end{bmatrix}$$

$$= \begin{bmatrix} R\varepsilon_{i0}^* + v_{i1} \\ v_{i2} \\ v_{i3} \\ \vdots \\ v_{i,T-1} \\ v_{iT} \end{bmatrix}$$

Denoting the $LT \times LT$ matrix in this expression by G^{-1}, we have

$$G = \begin{bmatrix} I_T & 0 & 0 & \cdots & 0 & 0 \\ R & I_T & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ R^{T-1} & R^{T-2} & R^{T-3} & \cdots & R & I_T \end{bmatrix}$$

Hence the conditional normal distribution of u_j^* has variance $G(I_T \otimes \Psi)G'$ and mean

$$\begin{bmatrix} \mu_1 + R\varepsilon_{i0}^* \\ \mu_2 + R^2\varepsilon_{i0}^* \\ \vdots \\ \mu_T + R^T\varepsilon_{i0}^* \end{bmatrix}.$$

The truncations of this distribution are linear and are given by (B.5). Hence the distribution of U_{ij}^* for given i and t, conditional on $U_{isl}^*(s \neq t \text{ or } j \neq l)$, ε_{i0}^*, and all the parameters of the models is truncated univariate normal. Therefore, $U_{11}^*, \ldots, U_{1T}^*, U_{21}^*, \ldots, U_{2T}^*, \ldots, U_{n1}^*, \ldots, U_{nT}^*$ may easily be drawn in succession, the drawn values replacing the old ones at each step. Details of this procedure are set out in Geweke (1991).

Step 2: Drawing ε_{i0}^* $(i = 1, \ldots, n)$. The kernel density of the conditional distribution of the ε_{i0}^* is the product of (B.1) and (B.4). These expressions reflect the assumption that the process $\{\varepsilon_{it}^*\}_{t=-\infty}^{\infty}$ is stationary. Since $\varepsilon_{it}^* = R\varepsilon_{i,t-1}^* + v_{it}$, the conditional distribution of ε_{i0}^* involves only ε_{i1}^*, R, and Ψ. It is indicated by the linear regression

$$\varepsilon_t = B\varepsilon_{t+1} + \zeta_t, \quad \operatorname{cov}(\varepsilon_{t+1}, \zeta_t) = 0$$

$B = [V_0(R, \Psi)]R[V_0(R, \Psi)]^{-1}$, and ζ_t has variance

$$V_0(R, \Psi) - BV_0(R, \Psi)B'.$$

Hence the conditional distribution of ε_{i0}^* is

$$\varepsilon_{i0}^* \sim N[B\varepsilon_{i1}^*, V_0(R, \Psi) - BV_0(R, \Psi)B'].$$

Step 3: Drawing ρ. The kernel density of the conditional distribution of $\rho = (\rho_1, \ldots, \rho_L)'$ is the product of (B.1), (B.4), and the kernel density of the truncated normal prior distribution of ρ. Expression (B.1), read as a function of ρ, is the kernel density of a multivariate normal distribution with precision (inverse variance)

$$H_d = \begin{bmatrix} \psi^{11}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{i1,t-1}^{*2} & \cdots & \psi^{1L}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{i1,t-1}^{*}\varepsilon_{iL,t-1}^{*} \\ \vdots & \ddots & \vdots \\ \psi^{L1}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{iL,t-1}^{*}\varepsilon_{i1,t-1}^{*} & \cdots & \psi^{LL}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{iL,t-1}^{*2} \end{bmatrix}$$

where $\Psi^{jk} \equiv [\Psi^{-1}]_{jk}$, and mean $H_d v_d$, where

$$v_j = \begin{bmatrix} \sum_{j=1}^{L}\psi^{1j}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{i1,t-1}^{*}\varepsilon_{ij,t-1}^{*} \\ \sum_{j=1}^{L}\psi^{Lj}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{iL,t-1}^{*}\varepsilon_{ij,t-1}^{*} \end{bmatrix}.$$

Let $V_\rho = \operatorname{diag}(\sigma_{\rho_1}^2, \ldots, \sigma_{\rho_L}^2)$. Then the distribution corresponding to the kernel density that is the product of (B.1) and the prior density for ρ is

$$N[H_d v_d + V_\rho^{-1}\rho, \quad (H_j + V_\rho^{-1})^{-1}] \tag{B.6}$$

truncated to the unit hypercube in R^L. The conditional distribution's kernel density is the product of the kernel density of this distribution and (B.4). Hence drawings from the full conditional distribution for ρ may be made by drawing from (B.6) and then using an acceptance step for the unit hypercube and (B.4). This may be done efficiently by noting that (B.4) is bounded above by

$|(1/N)S_{\varepsilon_0}|^{-n/2}\exp(-NL/2)$, where $S_{\varepsilon_0} \equiv \sum_{i=1}^{N} \varepsilon_{i0}\varepsilon_{i0}'$, because $(1/N)S_{\varepsilon_0}$ is the unconstrained conditional maximum likelihood estimate of $V_0 = \mathrm{var}(\varepsilon_{i0})$. Thus, drawigs for ρ are made from (B.4), rejected if $\rho_j < 0$ or $\rho_j \geqslant 1$ for any j, and then accepted with probability

$$|V_0(R,\Psi)|^{-N/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\,S_{\varepsilon_0}V_0(R,\Psi)^{-1}\right\}\Big/\left|\frac{1}{N}S_{\varepsilon_0}\right|^{-N/2}\exp\left(-\frac{NL}{2}\right). \tag{B.7}$$

The acceptance step is motivated by the similar procedures of Marriott et al. (1995) for stationary time series. The computation of (B.7) is trivial, and the fact that ε_{i0}^* is a synthetically drawn latent variable prevents acceptance probabilities from becoming impractically small.

Step 4: Drawing Ψ. The kernel density of the conditional distribution of Ψ is the product of (B.1), (B.4), and the kernel density of the inverted Wishart prior distribution of Ψ. The prior and (B.1) imply

$$\Psi^{-1} \sim W\left[S + \sum_{i=1}^{N}\sum_{t=1}^{T}(\varepsilon_{it}^* - R\varepsilon_{i,t-1}^*)(\varepsilon_{it}^* - R\varepsilon_{i,t-1}^*)', \nu + N\right].$$

The effect of (B.4) is then accommodated through an acceptance step just as it was in the drawing of ρ.

Step 5: Drawing $\tilde{\beta}_j^*(j = 1, \ldots, L)$. The kernel density of the conditional distribution of $\tilde{\beta}^* = (\tilde{\beta}_1^{*\prime}, \ldots, \tilde{\beta}_L^{*\prime})'$ is the product of (B.2) and the kernel density of the normal prior distribution of β. Since the model imposes no cross-equation constraints on the $\tilde{\beta}_j^*$ and the priors of the $\tilde{\beta}_j^*$ are independent, the conditional distribution of each $\tilde{\beta}_j^*$ has a simple form. Expression (B.2) as a function of $\tilde{\beta}_j^*$ is the kernel density of a multivariate normal distribution with precision

$$\psi^{jj}\sum_{i=1}^{N}\sum_{t=1}^{T}\dot{\tilde{X}}_{ijt}^*\dot{\tilde{X}}_{ijt}^{*\prime}$$

and mean

$$\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\dot{\tilde{X}}_{ijt}^*\dot{\tilde{X}}_{ijt}^{*\prime}\right]^{-1}\sum_{l=1}^{L}(\psi^{jl'}\psi^{jj})\sum_{i=1}^{N}\sum_{t=1}^{T}\dot{\tilde{X}}_{ijt}^* w_{ilt}^{(j)\prime},$$

where

$$w_{ilt}^{(j)} = \dot{U}_{ilt}^* - \dot{\tilde{X}}_{ilt}^{*\prime}\beta_l - Z_{ilt}'\gamma(l \neq j)$$

and

$$w_{ijt}^{(j)} = \dot{U}_{ijt}^* - Z_{ijt}'\gamma.$$

This mean and precision may then be combined with the prior mean and precision in the usual way to form the conditional, normal posterior distribution from which it is simple to make drawings.

Step 6: Drawing γ. The kernel density of the conditional distribution of γ is the product of (B.2) and the kernel density of the normal prior distribution of γ. Expression (B.2) as a function of γ is the kernel density of a multivariate normal distribution with precision

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{l=1}^{L}\sum_{j=1}^{L}\psi^{jl}\dot{Z}_{ijt}^{*}\dot{Z}_{ilt}^{*\prime}$$

and mean

$$\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{l=1}^{L}\sum_{j=1}^{L}\psi^{jl}\dot{Z}_{ijt}^{*}\dot{Z}_{ilt}^{*\prime}\right]^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{l=1}^{L}\sum_{j=1}^{L}\psi^{jl}\dot{Z}_{ijt}^{*}(\dot{U}_{ijt}^{*}-\dot{X}_{ijt}^{*}\bar{\beta}_{j}^{*}).$$

The mean and precision may then be combined with the prior mean and precision in the usual way to form the conditional, normal posterior distribution from which it is simple to make drawings.

From the structure of (B.2) and the prior distributions for the $\bar{\beta}_j^*$ and γ it is clear that the joint conditional posterior distribution of $(\bar{\beta}^{*\prime}, \gamma')'$ is normal, and therefore the $J+1$ drawings in steps 5 and 6 could be combined into one. This requires the solution of a much larger set of linear equations, each iteration of which results in greater execution time. On the other hand, the use of $J+1$ drawings rather than one introduces additional serial correlation into the Gibbs sampler. In the applications undertaken here, the choice is not important, because over 95% of execution time is devoted to drawing the U_{it}^* and ε_{i0}^*, and this step is the source of most serial correlation in the Gibbs sampler.

References

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

Elrod, T., Keane, M., 1995. A factor-analytic probit model for estimating market structure in panel data. Journal of Marketing Research 32, 1–16.

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

Geweke, J., 1991. Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints. In: Computer Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface, Alexandria, VA: American Statistical Association, pp. 571–578.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Berger, J.O., Bernardo, J.M., Dawid, A.P., Smith, A.F.M. (Eds.)

Proceedings of the Fourth Valencia International Meeting on Bayesian Statistics. Oxford University Press, Oxford, pp. 169–94.

Geweke, J., Keane, M., Runkle, D., 1994. Alternative computational approaches to statistical inference in the multinomial probit model. Review of Economics and Statistics 76, 609–632.

Gourieroux, C., Monfort, A., 1993. Simulation based inference: a survey with special reference to panel data models. Journal of Econometrics 59, 5–33.

Hajivassiliou, V., McFadden, D., 1994. A method of simulated scores for the estimation of LDV models. Discussion Paper # 967. Cowles Foundation.

Hajivassiliou, V., McFadden, D., Ruud, P., 1996. Simulation of multivariate normal rectangle probabilities and their derivatives: theoretical and computational results. Journal of Econometrics 72, 85–134.

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

Keane, M., 1994a. A computationally practical simulation estimator for panel data. Econometrica 62 (1), 95–116.

Keane, M., 1994b. Modelling heterogeneity and state dependence in consumer choice behavior. Working Paper. University of Minnesota.

Lee, L. F., 1992. On efficiency of methods of simulated moments and maximum simulated likelihood estimation of discrete response models. Econometric Theory 8, 518–52.

Lee, L.F., 1995. Asymptotic bias in maximum simulated likelihood estimation of discrete choice models. Econometric Theory 11, 437–483.

Marriott, J., Ravishanker, N., Gelfand, A., Pai, J., 1995. Bayesian analysis of ARMA processes: Complete sampling based inference under exact likelihood. In: Berry, D., Geweke, J., Chaloner, K., (Eds.) Bayesian Statistics and Econometrics, Wiley, New York.

McCulloch, R., Rossi, P. E., 1994. An exact likelihood analysis of the multinomial probit model. Journal of Econometrics 64, 207–240.

McFadden, D., 1989. A method of simulated moments for estimation of discrete response models without numerical integration. Econometrica 57, 995–1026.

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

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