Federal Reserve Bank of Minneapolis
Research Department Staff Report 177

July 1994

Statistical Inference in the Multinomial Multiperiod Probit Model

John F. Geweke*

Federal Reserve Bank of Minneapolis
and University of Minnesota

Michael P. Keane*

Federal Reserve Bank of Minneapolis

David E. Runkle*

Federal Reserve Bank of Minneapolis

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 approaches to inference that circumvent the integration problem: Bayesian inference using Gibbs sampling and data augmentation to compute posterior moments, simulated maximum likelihood (SML) estimation using the GHK recursive probability simulator, and method of simulated moment (MSM) estimation using the GHK simulator. We perform a set of Monte-Carlo experiments to compare the performance of these approaches. Although all the methods perform reasonably well, some important differences emerge. The root mean square errors (RMSEs) of the SML parameter estimates around the data generating values exceed those of the MSM estimates by 21 percent on average, while the RMSEs of the MSM estimates exceed those of the posterior parameter means obtained via Gibbs sampling by 18 percent on average. While MSM produces a good agreement between empirical RMSEs and asymptotic standard errors, the RMSEs of the SML estimates exceed the asymptotic standard errors by 28 percent on average. Also, the SML estimates of serial correlation parameters exhibit significant downward bias.

*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.

I. 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 1994, Keane 1994b, McCulloch and Rossi 1994), choice of residential location (Hajivassiliou et al. 1993), 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 alternative approaches to simulation based inference for the MMP model in samples representative of these applications. The goal of the present paper is to provide such a comparison. First, we describe three simulation based 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 Bayesian inference based on Gibbs sampling and data augmentation. We perform a set of Monte-Carlo experiments to compare the performance of these approaches. 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.

Although all three approaches to inference perform reasonably well in our experiments, some important differences emerge. The root mean square errors (RMSEs) of the simulated maximum likelihood point estimates around the data generating values exceed those of the method of simulated moments estimates by 21 percent on average, while the RMSEs of the MSM estimates exceed those of the posterior means obtained via Bayesian inference by 18 percent on average. While MSM produces a good agreement between empirical RMSEs and asymptotic standard errors, the RMSEs of the SML estimates exceed the asymptotic standard errors by 28 percent on average. Also, the SML estimates of serial correlation parameters exhibit significant downward bias.

While this is the first systematic study of the performance of simulation based approaches to inference in the MMP model in representative samples, a number of authors have studied the single period case. McCulloch and Rossi (1994) studied Bayesian inference using Gibbs sampling and data augmentation to compute posterior moments in a single period multinomial probit model. Geweke, Keane, and Runkle (1994) compared four alternative methods of inference in the single period multinomial probit model: Bayesian inference using Gibbs sampling and data augmentation, MSM and SML using the GHK probability simulator, and MSM estimation based on a kernel smoothed frequency simulator. Only Keane (1994a) has studied simulation estimation in the multiperiod probit model, and he examined both MSM and SML estimation based on the GHK probability simulator. However, he only considered binomial probit models. Furthermore, he did not consider Bayesian inference and did not utilize an experimental design that allows evaluation of the influence of the data structure (for example, serial correlation of the exogenous variables) on the performance of the alternative methods.

In Section II we describe the MMP model. Section III takes up two simulation based approaches to classical inference in this model (SML and MSM) and the GHK method for simulation of choice probabilities upon which these approaches are based. In Section IV we describe the Bayesian approach to inference using Gibbs sampling and data augmentation to compute posterior moments. Section V lays out the experimental design of our Monte-Carlo study, Section VI presents the results, and Section VII concludes.

II. 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

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

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 unknown 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

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

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

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

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

$$(2.1)\quad U^*_{ijt} = (U_{ijt}-U_{iJt})(\sigma_{11}+\sigma_{JJ}-2\sigma_{1J})^{-1/2}$$
$$= [(X'_{ijt}\beta_j - X'_{iJt}\beta_J) + (\epsilon_{ijt}-\epsilon_{iJt})](\sigma_{11}+\sigma_{JJ}-2\sigma_{1J})^{-1/2}$$
$$= X^{*\prime}_{ijt}\beta^*_j + \epsilon^*_{ijt} \quad (j = 1,J;\ t = 1,T),$$

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 $\epsilon^*_{iJt} = 0$.) We further define

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

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

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

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

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

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

$$P(d_i) = P(U^*_{i,j_{it},t} > U^*_{ikt}\ \forall\ k \neq j_{it},\ t = 1,\ldots,T)$$
$$= P\left[\epsilon^*_{i,j_{it},t} - \epsilon^*_{ikt} > X^{*\prime}_{ikt}\beta^*_k - X^{*\prime}_{i,j_{it},t}\beta^*_{j_{it}}\ \forall\ k \neq j_{it},\quad (t = 1,\ldots,T)\right].$$

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 $T \cdot (J-1)$ 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 $2 \cdot (J-1)$. The goal of simulation based inference is to allow a richer covariance structure to be used.

In this paper we consider a particular special case of the model (2.1)–(2.4) in which the ϵ^*_{ijt} are stationary first-order autoregressive 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 Z^*_{ijt} set that varies across alternatives (which can be thought of as containing attributes of alternatives, such as price or quality), but for which the corresponding coefficient is restricted to be equal across alternatives.

These decisions are motivated by the 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 first-order autoregressive 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 (Keane 1990).

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

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

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

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

In matrix notation, the model is then

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

Finally, adopt the notational convention $u_{i0}^{*} = \epsilon_{i0}^{*}$, $\tilde{X}_{i0}^{*} = [0]$. The random utility component ϵ_{it}^{*} follows a conventional, stationary, first-order autoregressive process

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

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.

III. Classical Approaches to Inference

A. Simulation of Choice Sequence Probabilities

Classical approaches to inference in 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, McFadden, and Ruud (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, Keane, and Runkle (1994), in a Monte-

Carlo study of alternative approaches to simulation based inference 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 simulation based estimators in this paper. Here we provide a description of the GHK simulator applied to the simulation of choice probabilities on the MMP model just described. 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

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

(Notice that $\tilde{U}^j_{ijt} = 0$ and $\tilde{\epsilon}^j_{ijt} = 0$.) 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

$$\tilde{\epsilon}_{it}(-j) = (\tilde{\epsilon}^j_{i1t},\ldots,\tilde{\epsilon}^j_{i,j-1,t},\tilde{\epsilon}^j_{i,j+1,t},\ldots,\tilde{\epsilon}^j_{iJt})' \text{ and } \tilde{\epsilon}(d_i) = (\tilde{\epsilon}_{i1}(-j_{i1}),\ldots,\tilde{\epsilon}_{iT}(-j_{iT}))'$$

where d_i is the choice vector defined in (2.4). Thus $\tilde{\epsilon}(d_i) \sim \text{IIDN}(0,\tilde{\Sigma}(d_i))$, 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 $\tilde{\epsilon}(d_i) = \tilde{A}(d_i)\tilde{\eta}(d_i)$, 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 $\eta_{ijt} \sim \text{IIDN}(0,1)$ for all i,j,t.

Finally, define $U^j_{ikt}(\tilde{\eta}^\ell_{11},\ldots,\tilde{\eta}^\ell_{J,t-1},\tilde{\eta}^\ell_{1t},\ldots,\tilde{\eta}^\ell_{pt})$ 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* $(\tilde{\eta}^\ell_{11},\ldots,\tilde{\eta}^\ell_{J,t-1},\tilde{\eta}^\ell_{1t},\ldots,\tilde{\eta}^\ell_{pt})$. 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_{11}^{ℓ} s.t. $\tilde{U}_{11}^{j_1}(\eta_{11}^{\ell}) < 0$

⋮

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

(j_1) Skip $\eta_{j_1,1}^{\ell}$

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

⋮

(J) Draw η_{J1}^{ℓ} s.t. $\tilde{U}_{J1}^{j_1}(\eta_{11}^{\ell},\ldots,\eta_{j_1-1,1}^{\ell},\eta_{j_1+1,1}^{\ell},\ldots,\eta_{J1}^{\ell}) < 0$

⋮

Period t:

Step:

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

⋮

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

(j_t) Skip $\eta_{j_t t}^{\ell}$

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

⋮

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

and finally, construct:

$$\hat{P}_{GHK}(d_1,\ldots,d_t|\beta^*,\Sigma^*,X^*)$$

$$= \frac{1}{M}\sum_{\ell=1}^{M} P(\tilde{U}_{11}^{j_1} < 0)\prod_{k=2}^{j_1-1} P\left[\tilde{U}_{k1}^{j_1}\left(\eta_{11}^{\ell},\ldots,\eta_{k-1,1}^{\ell}\right) < 0\right] \cdot P\left[\tilde{U}_{j_1+1,1}^{j_1}\left(\eta_{11}^{\ell},\ldots,\eta_{j_1-1,1}^{\ell}\right) < 0\right]$$

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

$$\cdot \ \ldots \ \cdot P\left(\tilde{U}_{1t}^{j_t}\left(\tilde{\eta}_{11}^{\ell},\ldots,\tilde{\eta}_{Jt-1}^{\ell}\right) < 0\right)\prod_{k=2}^{j_t-1} P\left[\tilde{U}_{kt}^{j_t}\left(\tilde{\eta}_{11}^{\ell},\ldots,\tilde{\eta}_{Jt-1}^{\ell},\eta_{1t}^{\ell},\ldots,\eta_{k-1,t}^{\ell}\right) < 0\right]$$

$$P\left[\tilde{U}_{j_t+1,t}^{j_t}\left(\tilde{\eta}_{11}^{\ell},\ldots,\tilde{\eta}_{J_1,t-1}^{\ell},\tilde{\eta}_{1t}^{\ell},\ldots,\tilde{\eta}_{j_t-1,t}^{\ell}\right) < 0\right]$$

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

B. 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:

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

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

A direct application of the MSM estimator developed in McFadden (1989) to the MMP model would involve indexing all possible choice sequences $s = 1, \ldots, J^T$ and defining the choice indicators $d_{is} = 1$ if i chooses sequence s and 0 otherwise. Then form the MSM estimator by solving the moment conditions:

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

This MSM estimator is consistent for fixed M. This direct approach is not feasible because of the computational burden involved in simulating the 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:

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

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

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

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 superior small sample properties to SML, especially when serial correlation is strong. (This study was limited to binomial probit models. We provide evidence for the trinomial probit case, and Keane's finding will be strengthened.)

IV. Bayesian Inference Using the Gibbs Sampler

Bayesian inference using the Gibbs sampler (Gelfand and Smith 1990) and data augmentation (Tanner and Wong 1987) has been applied to the multinomial multiperiod probit model by Allenby and Rossi (1993). The approach taken here follows Allenby and Rossi but differs in 3 respects: Here, all priors are proper whereas Allenby and Rossi used improper priors for β^*; 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. In a Bayesian approach, the latent utilities U^*_{ijt} and the coefficient vectors $\beta^*_j (j = 1,\ldots,J-1)$, γ, and Σ^* enter the model in symmetric fashion, having a joint posterior distribution conditional on the data. The computation of a posterior moment involves integration over both the latent utilities and the parameters. The Gibbs sampling-data augmentation algorithm executes this computation by constructing an artificial sample from the joint posterior distribution and then forming sample means of the functions whose posterior moments are of interest.

To provide an overview of the algorithm in generic notation, suppose a model specifies a probability density function $f_1(Y|\tilde{Y},\theta)$ for observable random vector Y conditional on the latent random vector $\tilde{Y}$ and parameter vector θ. Suppose the model also specifies a probability density function $f_2(\tilde{Y}|\theta)$ for $\tilde{Y}$. Having observed Y, we know the likelihood function is

$$L(Y|\theta) = \int_{\tilde{Y}} f_1(Y|\tilde{Y},\theta) f_2(\tilde{Y}|\theta) d\tilde{Y}.$$

The explicit integration in this expression is often quite difficult, and consequently maximum likelihood or other optimization estimates are nontrivial. (The SML estimator described above is an example.) Given a prior probability density for θ with kernel $f_0(\theta)$, the posterior probability density for θ is

$$p(\theta|Y) = \frac{\int_{\tilde{Y}} f_0(\theta) f_1(Y|\tilde{Y},\theta) f_2(\tilde{Y}|\theta) d\tilde{Y}}{\int_{\Theta}\int_{\tilde{Y}} f_0(\theta) f_1(Y|\tilde{Y},\theta) f_2(\tilde{Y}|\theta) d\tilde{Y} d\theta}$$

and the posterior expectation of a function of interest $g(\theta,\tilde{Y})$ is

$$\begin{aligned} E[g(\tilde{Y},\theta)] &= \frac{\int_{\Theta}\int_{\tilde{Y}} g(\theta,\tilde{Y}) f_0(\theta) f_1(Y|\tilde{Y},\theta) f_2(\tilde{Y}|\theta) d\tilde{Y}}{\int_{\Theta}\int_{\tilde{Y}} f_0(\theta) f_1(Y|\tilde{Y},\theta) f_2(\tilde{Y}|\theta) d\tilde{Y} d\theta} \\ &= \int_{\Theta} g(\theta,\tilde{Y}) p(\theta,\tilde{Y}|Y) d\tilde{Y} d\theta, \end{aligned}$$

where $p(\theta,\tilde{Y}|Y)$ is the joint posterior density function for θ and $\tilde{Y}$.

Suppose there is a partition of the parameter vector θ into m subvectors, $\theta' = (\theta'_{(1)},\ldots,\theta'_{(m)})$, such that the conditional posterior density functions

$$p(\theta_{(i)}|\theta_{(j)}, j \neq i,\tilde{Y},Y)$$

are of sufficiently simple form that synthetic random vectors $\tilde{\theta}_{(i)}$ from the corresponding distribution can be drawn. Suppose further that the conditional posterior density function

$$p(\tilde{Y}|\theta,Y)$$

is of a form that the generalized signal extraction problem of drawing synthetic random vectors from the corresponding distribution has a practical solution. (An example will be provided below.) The Gibbs sampling-data augmentation algorithm proceeds beginning from an initial value $(\theta^{(0)},\tilde{Y}^{(0)})$ in the support of $p(\theta,\tilde{Y}|Y)$, drawing in rotation $\tilde{Y}$, $\theta_{(1)}$, ..., $\theta_{(m)}$ from the appropriate conditional distribution and replacing the previous value with the one just drawn.

To describe the implementation of this method for the multinomial multiperiod probit 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

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

The values of the β^*_j change accordingly, as does $\Psi = \mathrm{var}(\epsilon^*_{it}-R\epsilon^*_{i,t-1})$. The diagonal 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 distribution was used in all cases:

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

$$\gamma \sim N(\underline{\gamma}, V_\gamma)$$

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

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

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

where W denotes the Wishart distribution; and the prior distributions of these 2J components are mutually independent. These proper prior distributions identify the $\tilde{\beta}_j^*$, γ, and Ψ which otherwise would be identified only up to a scale factor. Then $\tilde{\beta}_j^*(\sigma_{11}^*)^{-1/2}$ is comparable with $\tilde{\beta}_j^*$ of Section II $(j = 1,\ldots,L)$; $\gamma(\sigma_{11}^*)^{-1/2}$ is comparable with γ of Section II; and $(\sigma_{11}^*)^{-1/2}\,\Psi$ is comparable with Ψ of Section II.

Equivalent expressions for the probability density kernel for the U_{it}^* are

$$(4.1) \quad |\Psi|^{-NT/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}(\epsilon_{it}^* - R\epsilon_{i,t-1}^*)'\Psi^{-1}(\epsilon_{it}^* - R\epsilon_{i,t-1}^*)\right\},$$

where $\epsilon_{it}^* = U_{it}^* - X_{it}^*\tilde{\beta}^* - Z_{it}^*\gamma$; or

$$(4.2) \quad |\Psi|^{-NT/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\sum_{i=1}^{T}(\dot{U}_{it}^* - \dot{X}_{it}^*\tilde{\beta}^* - \dot{Z}_{it}^*\gamma)'\Psi^{-1}(\dot{U}_{it}^* - \dot{X}_{it}^*\tilde{\beta}^* - \dot{Z}_{it}^*\gamma)\right\},$$

where

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

or

$$(4.3)\quad |\Psi|^{-NT/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[U_{it}^{*}-RU_{i,t-1}^{*}-(\mu_{it}-R\mu_{i,t-1})\right]\Psi^{-1}\left[U_{it}^{*}-RU_{i,t-1}^{*}-(\mu_{it}-R\mu_{i,t-1})\right]\right\},$$

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

$$(4.4)\quad |V_0(R,\Psi)|^{-N/2}\exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\epsilon_{i0}^{*\prime}[V_0(R,\Psi)]^{-1}\epsilon_{i0}^{*}\right\},$$

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 (4.1) or (4.2) or (4.3); (4.4); the density kernels for the prior distributions; and

$$(4.5)\quad \prod_{i=1}^{N}\prod_{t=1}^{T} H(U_{it}^{*},d_{iT}),$$

where H is an indicator function for consistency of orderings and signs of the $U_{ijt}^{*}(j = 1,\ldots,L)$ with the observed choice vectors d_{it}. Using notation defined in Section II, 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 (4.3) and (4.5). The conditional distribution of $U_{i}^{*} = (U_{i1}^{*\prime},\ldots,U_{iT}^{*\prime})'$ is truncated normal. The conditional normal distribution is given by the relations

$$U^*_{i1} - \mu_{i1} - R\epsilon^*_{i0} = v_{i1}$$

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

$$\vdots$$

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

from (4.3), which imply

$$\begin{bmatrix} I_T & 0 & 0 & \cdots & 0 & 0 & 0 \\ -R & I_T & 0 & \cdots & 0 & 0 & 0 \\ 0 & -R & I_T & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & & & \\ 0 & 0 & 0 & \cdots & -R & I_T & 0 \\ 0 & 0 & 0 & \cdots & 0 & -R & I_T \end{bmatrix} \begin{bmatrix} U^*_{it} - \mu_{i1} \\ U^*_{i2} - \mu_{i2} \\ U^*_{i3} - \mu_{i3} \\ \vdots \\ U_{i,T-1} - \mu_{i,T-1} \\ U^*_{iT} - \mu_{iT} \end{bmatrix} = \begin{bmatrix} R\epsilon^*_{i0} + v_{i1} \\ v_{i2} \\ v_{i3} \\ \vdots \\ v_{i,T-1} \\ v_{iT} \end{bmatrix}.$$

Denoting the LT × 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

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

The truncations of this distribution are linear and are given by (4.5). Hence the distribution of U^*_{itj} for given i and t, conditional on U^*_{isf}($s \neq t$ or $j \neq \ell$), ϵ^*_{i0}, and all the parameters of the models is truncated univariate normal. Therefore, U^*_{11}, ..., U^*_{1T}, U^*_{21}, ..., U^*_{2T}, ..., U^*_{n1}, ..., 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 $\epsilon^*_{i0}(i = 1,\ldots,n)$. The kernel density of the conditional distribution of the ϵ^*_{i0} is the product of (4.1) and (4.4). These expressions reflect the assumption that the process $\{\epsilon^*_{it}\}_{t=-\infty}^{\infty}$ is stationary. Since $\epsilon^*_{it} = R\epsilon^*_{i,t-1} + v_{it}$, the conditional distribution of ϵ^*_{i0} involves only ϵ^*_{i1}, R, and Ψ. It is indicated by the linear regression

$$\epsilon_t = B\epsilon_{t+1} + \zeta_t, \quad \mathrm{cov}(\epsilon_{t+1},\zeta_t) = 0$$

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

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

Hence the conditional distribution of ϵ^*_{i0} is

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

Step 3. Drawing ρ. The kernel density of the conditional distribution of $\rho = (\rho_1,\ldots,\rho_L)'$ is the product of (4.1), (4.4), and the kernel density of the truncated normal prior distribution of ρ. Expression (4.1), read as a function of ρ, is the kernel density of a multivariate normal distribution with precision (inverse variance)

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

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

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

Let $\varrho' = (\varrho_1',\ldots,\varrho_L')$ and $V_\rho = \text{diag}(\sigma^2_{\rho_1},\ldots,\sigma^2_{\rho_L})$. Then the distribution corresponding to the kernel density that is the product of (4.1) and the prior density for ρ is

$$(4.6) \quad N[H_d v_d + V_\rho^{-1}\varrho, \quad (H_j + V_\rho^{-1})^{-1}]$$

truncated to the unit hypercube in $\mathbf{R}^L$. The conditional distribution's kernel density is the product of the kernel density of this distribution and (4.4). Hence drawings from the full conditional distribution for ρ may be made by drawing from (4.6) and then using an acceptance step for the unit hypercube and (4.4). This may be done efficiently by noting that (4.4) is bounded above by $|(1/N)S_{\epsilon_0}|^{-n/2}\exp(-NL/2)$, where $S_{\epsilon_0} \equiv \sum_{i=1}^{N}\epsilon_{i0}\epsilon_{i0}'$, because $(1/N)S_{\epsilon_0}$ is the unconstrained conditional maximum likelihood estimate of $V_0 = \text{var}(\epsilon_{i0})$. Thus, drawings for ρ are made from (4.4), rejected if $\rho_j < 0$ or $\rho_j \geq 1$ for any j, and then accepted with probability

$$(4.7) \quad |V_0(R,\Psi)|^{-N/2}\exp\left\{-\frac{1}{2}\,\text{tr}S_{\epsilon_0}V_0(R,\Psi)^{-1}\right\}\Big/|\frac{1}{N}\,S_{\epsilon_0}|^{-N/2}\exp\left[-\frac{NL}{2}\right].$$

The acceptance step is motivated by the similar procedures of Marriott et al. (forthcoming) for stationary time series. The computation of (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 (4.1), (4.4), and the kernel density of the inverted Wishart prior distribution of Ψ. The prior and (4.1) imply

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

The effect of (4.4) 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 (4.2) 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 (4.2) as a function of $\tilde{\beta}_j^{*}$ is the kernel density of a multivariate normal distribution with precision

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

and mean

$$\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\dot{X}_{ijt}^{*}\dot{X}_{ijt}^{*\prime}\right]^{-1}\sum_{\ell=1}^{L}(\psi^{i\ell\prime}\psi^{ij})\sum_{i=1}^{N}\sum_{t=1}^{T}\dot{X}_{ijt}^{*}w_{i\ell t}^{(j)\prime},$$

where

$$w_{i\ell t}^{(j)} = \dot{U}_{i\ell t}^{*} - \dot{X}_{i\ell t}^{*\prime}\beta_{\ell} - Z_{i\ell t}^{\prime}\gamma(\ell \neq j)$$

and

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

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 (4.2) and the kernel density of the normal prior distribution of γ. Expression (4.2) as a function of γ is the kernel density of a multivariate normal distribution with precision

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{\ell=1}^{L}\sum_{j=1}^{L}\psi^{j\ell}\tilde{Z}^{*}_{ijt}\tilde{Z}^{*\prime}_{i\ell t}$$

and mean

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

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 (4.2) and the prior distributions for the $\tilde{\beta}^{*}_{j}$ and γ it is clear that the joint conditional posterior distribution of $(\tilde{\beta}^{*\prime},\gamma')'$ 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 percent 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.

The same prior distributions were employed throughout the experiment:

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

The effect of the prior distribution for Ψ is to center the ψ_{jj}, which otherwise would be identified only up to a scale factor, about 1.0. This in turn would induce a proper posterior distribution for the $\tilde{\beta}^{*}_{j}$ and γ, even if the priors for these coefficients were flat and improper. (This technique was

introduced by McCulloch and Rossi (1994) in the one-period multinomial probit model.) The posterior distribution of these parameters induces a posterior distribution on $\Sigma^* = \text{var}(\epsilon^*_{it})$, with $\sigma^*_{jk} = \psi_{jk}/(1-\rho_j\rho_k)$. In the context of the normalization set forth in Section II, parameters of interest are $\beta^*_j(\sigma^*_{11})^{-1/2}$(j = 1,...,L); $\gamma(\sigma^*_{11})^{-1/2}$; ρ_j for j = 1,...,L; and the lower triangular matrix A*: $A^*A^{*\prime} = (\sigma^*_{11})^{-1}\Psi$. To make drawings from the posterior distribution of these functions it is necessary only to transform the drawings of the β^*_j, γ, ρ_j, and Ψ.

The first 200 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. The next m = 5,000 iterations were retained, creating the posterior sample

$$\{\tilde{\beta}^{*(j)},\gamma^{(j)},\rho^{(j)},\Psi^{(j)}\}_{j=1}^{m}.$$

The posterior mean of a function of interest

$$\bar{g} = E[g(\tilde{\beta}^*,\gamma,\rho,\Psi)]$$

is approximated by the corresponding posterior sample moment,

$$\bar{g}_m = m^{-1}\sum_{j=1}^{m} g(\tilde{\beta}^{*(j)},\gamma^{(j)},\rho^{(j)},\Psi^{(j)}).$$

The standard error of this Monte-Carlo approximation to $\bar{g}$ was assessed as described in Gewèke (1992). Typically, this standard error was less than 10 percent of the posterior standard deviation for the experiments undertaken.

V. Experimental Design

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

$$U_{i1t} = 0.5 + 1 \cdot X_{it} + 1 \cdot Z_{it1} + \epsilon_{i1t}$$

$$U_{i2t} = -1.2 + 1 \cdot X_{it} + 1 \cdot Z_{it2} + \epsilon_{i2t}$$

$$U_{i3t} = 0.0.$$

The random shocks to utility evolve according to

$$\epsilon_{i1t} = \rho_1 \epsilon_{i1,t-1} + v_{i1t}$$

$$\epsilon_{i2t} = \rho_2 \epsilon_{i2,t-1} + v_{i2t}$$

$$\begin{bmatrix} v_{i1t} \\ v_{i2t} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ a_{21} & \sqrt{1-a_{21}^2} \end{bmatrix} \begin{bmatrix} \eta_{i1t} \\ \eta_{i2t} \end{bmatrix}$$

with $\eta_{it} \sim$ IIDN(0,I(1$-\rho_1^2$)). In the notation of Section II,

$$\beta_1 = (0.5,1,1)', \quad \beta_2 = (-1.2,1,1)'$$

$$X_{i1t1} = X_{i2t1} \equiv X_{it}, \quad X_{i1t2} \equiv Z_{i1t}, \quad X_{i2t2} \equiv Z_{i2t}$$

$$\beta_{13} = \beta_{23} \equiv \gamma,$$

where $X_{ijt\ell}$ refers to the ℓth element of the X_{ijt} vector. The regressors are constructed as follows:

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

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

with

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

In our experiment we use a 3 × 2 × 2 full factorial design,

$$\rho = 0.50 \text{ or } 0.80; \quad a_{21} = 0.50 \text{ or } 0.80; \quad \phi^2 = 0 \text{ or } 0.50 \text{ or } 0.80.$$

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.

VI. Results

The results of the 12 Monte-Carlo experiments are reported in Tables 1–12. For the classical methods 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, RMSE; and 3) the mean of the asymptotic standard errors across the 20 replications, $\overline{\text{ASE}}$. For Bayesian inference we report 1) the mean of the posterior means across the 20 replications, $\bar{\hat{\theta}}$; 2) the RMSE of the posterior means around the data generating values, $\overline{\text{RMSE}}$; and 3) the mean of the posterior standard deviation across the 20 replications, $\overline{\text{PSD}}$.

In the remainder of this section, we compare the performance of the different methods in each experiment, focussing primarily on the RMSEs as the criterion of performance. In order to make possible comparisons between the classical and Bayesian methods, we treat the empirical RMSEs of the classical point estimates around the data generating parameter values as comparable to the empirical RMSEs of the Bayesian posterior means around those values. Also of interest are comparisons of empirical RMSEs with ASEs for the classical point estimates. In this context, we also compare empirical RMSEs of the Bayesian posterior means with the corresponding PSDs.

In Table 1 we consider the model with $\rho = 0.50$, $a_{21} = 0.50$, and $\phi^2 = 0$. This is the case of low serial correlation and low cross correlation of the random elements of utility combined with no serial correlation in the exogenous variables. For Bayesian inference, the mean across the 20 replications of the posterior means are close to the data generating values for all 9 model parameters. This is also true of the MSM point estimates. SML, on the other hand, while producing point estimates close to the data generating values for most model parameters, exhibits severe bias in estimating the degree of serial correlation. 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 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 Bayesian inference, marginally significant biases are found only for ρ_2 and β_{12} if the Bayesian posterior means are viewed as estimates of the data generating parameter values.

In comparison of RMSEs, Bayesian inference has an edge over the classical methods. The RMSEs of the posterior means around the data generating values are smaller than the corresponding RMSEs of the classical point estimates around the data generating values for 6 of the 9 model parameters. Exceptions are β_{12} and β_{22} for which the RMSEs of the MSM point estimates are smallest, and β_{22} for which the RMSE of the SML point estimates is smallest. Another clear pattern is that for MSM the empirical RMSEs and the mean asymptotic standard errors are in close agreement for most model parameters. But for SML, the mean asymptotic standard errors are substantially below the empirical RMSEs for the covariance matrix parameters. Interestingly, the mean asymptotic standard errors for MSM and the mean posterior standard deviations from Bayesian inference are in very close agreement. Given that the RMSEs for Bayesian inference are generally lower than for MSM, this also means that the mean posterior standard deviations obtained via Bayesian inference are generally well above the RMSEs of the posterior means about the data generating values.

Rather than describing Tables 2–12 with the same level of detail devoted to Table 1, we instead point out certain broad patterns that emerge. As we move across Tables 1 to 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 point estimates improve relative to those for the other methods as ϕ^2 increases.

In Tables 4–6 the degree of serial correlation in the random components of utility is increased (with ρ_1 and ρ_2 being set at 0.80). Again, as we move across Tables 4 to 6, the 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 Bayesian inference, 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, there is a pattern of rising RMSEs as ϕ^2 increases. This effect seems to be greater in Tables 4–6. Thus, increased serial correlation of the errors seems to accentuate the positive impact of increasing serial correlation in covariates on RMSEs. The ρ are again an exception to this pattern, as the RMSEs of the ρ again appear to fall as ϕ^2 increases.

In Tables 7–9 the degree of serial correlation in the random components of utility 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 7 of 9 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 4 of 9 parameters. These improvements in relative performance for SML and MSM do not extend to other cases.

In Tables 10–12 both the degree of serial and cross correlations in the errors is set at the "high" level ($\rho_1 = \rho_2 = 0.80$, $a_{12} = 0.80$). Again, in moving across Tables 10 to 12, the degree of serial correlation in the regressors is increased. The RMSEs for all model parameters except the ρ tend to be higher than in Tables 1–3, and this effect is most pronounced for MSM. For the ρ the RMSEs are clearly lower than in Tables 1–3. As we move across Tables 10–12, it is again obvious that the RMSEs for all model parameters except the ρ tend to rise as serial correlation in the

covariates increases. The RMSEs for the ρ tend to fall. There is relative improvement in the performance of MSM as ϕ^2 increases.

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 all methods.

Averaging across all 12 models and all 9 parameters, the ratio of RMSE for SML to that for Bayes is 1.39. The ratio of RMSE for MSM to that for Bayes is 1.18. In Table 13 we present for each of the three methods the means of the RMSEs for each model parameter across all 12 experiments. We also report the mean bias and mean ASE or PSD. The Bayesian posterior means have the lowest (or tied for lowest) RMSEs about the data generating parameter values for 7 of the 9 model parameters. The exceptions are β_{12} and β_{22}, the slope coefficients on the variable X_{it} in the utility functions for alternatives 1 and 2. For both of these parameters, the RMSEs of the SML point estimates are the smallest. Nevertheless, for all parameters except ρ_1 and ρ_2, the RMSEs produced by all three methods are typically within a range of 10 to 20 percent of one another. The principal weakness of SML appears in the very large RMSEs of the SML point estimates for ρ_1 and ρ_2, which are 80 to 150 percent greater than those of MSM or Bayesian inference.

In Table 13 we also see that agreement between RMSEs and ASEs is quite close for MSM. For SML, the ASEs are in the range of 40 percent to 60 percent below the RMSEs for all four of the error covariance parameters. There is some tendency for the PSDs to be below the RMSEs of the posterior means for Bayesian inference. The grand means of the PSDs or ASEs are 0.060 for

Bayesian inference, 0.065 for MSM, and 0.049 for SML. Thus, the RMSEs of the SML point estimates exceed the asymptotic standard errors by 38 percent on average.

In Table 14 we present a regression that summarizes the relative performance of the alternative methods. The dependent variable in this regression is the log RMSE of the point estimates or posterior means for a parameter in one of the experiments. The right hand side variables are dummies for parameter, method of inference, and different levels of the treatments (that is, degree of cross sectional correlation in the errors, degree of serial correlation in the errors, 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 Bayesian inference for γ in the model with $\rho = 0.50$, $a_{21} = 0.50$, 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 percent greater than the RMSEs of the Bayesian posterior means. For SML the corresponding estimate is roughly 4 percent. Turning to the coefficients on the model parameter dummies, we see that the cross correlation parameters a_{21} and a_{22} are much less precisely estimated than most other model parameters. Also ρ_2, β_{21}, and β_{22} are much less precisely estimated than ρ_1, β_{11}, and β_{12}. This presumably occurs because the fraction of the population that chooses alternative 1 is substantially greater than the fraction that chooses alternative 2.

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. Note that all the parameter with MSM-GHK interactions are insignificant except for those involving β_{12} and β_{22}, the coefficients on X_{it} in the alternative 1 and 2 utility functions. These are -0.181 and -0.195, respectively, which, when combined with the MSM main effect of 0.165 reported above, indicates that MSM produces slightly smaller RMSEs than Bayesian inference

for these parameters (as was already apparent in Table 13). The parameter with SML-GHK interactions produce 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 percent level. Thus, the RMSEs of the SML point estimates for the serial correlation parameters are roughly 76 percent and 90 percent greater than those of the Bayesian 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.

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 serial correlation in the errors from 0.50 to 0.80 causes the RMSEs for Bayesian inference to rise by roughly 12 percent 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 Bayesian inference to fall by roughly 30 percent.

The main effects of the $\phi^2 = 0.50$ and $\phi^2 = 0.80$ treatments are 0.185 and 0.468. Thus, raising the correlation across time periods in the regressors from 0 to 0.50 or from 0 to 0.80 causes the RMSEs for Bayesian inference to rise by roughly 18 percent and 46 percent for all parameters other than the ρ. However, the interactions of the $\phi^2 = 0.50$ and $\phi^2 = 0.80$ dummies with the ρ parameter dummy have coefficient estimates of -0.383 and -0.530. These imply that raising the correlation across time periods in the regressors from 0 to 0.50 or from 0 to 0.80 reduces the RMSEs for the ρ parameters by roughly 20 percent and 6 percent, respectively. Thus, we see that

increasing either serial correlation in the random parts of utility or in the covariates causes the serial correlation parameters to be more precisely identified.

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 percent and the drops in RMSEs for the ρ parameters when serial correlation is strong are only about 13–14 percent. Thus, the performance of Bayesian inference relative to classical methods improves significantly as serial correlation is increased. This result may be surprising, because the serial correlation in the Gibbs draws for the latent variables is increasing when the serial correlation in the random component of utility is increased. However, the accuracy of the GHK method for simulation of choice sequence probabilities will also decrease.

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 -0.148, which indicates that the performance of SML relative to Bayesian inference 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.

Interestingly, increasing the cross correlation parameter a_{21} from 0.50 to 0.80 has little effect on the precision with which Bayesian inference uncovers the parameters of the data generating process. However, the significant coefficients of -0.082 and -0.129 on the interactions of the a_{21} dummy with MSM-GHK and SML-GHK indicate that the relative performance of MSM and SML improves as cross correlation is increased.

VII. Conclusion

We find that Bayesian inference using Gibbs sampling and data augmentation to compute posterior moments, MSM estimation using the GHK probability simulator, and SML using the GHK probability simulator all perform reasonably well in point estimation of parameters of the data generating process in a three alternative 10-period multinomial multiperiod probit model. However, some important patterns emerge regarding relative performance of the methods—both in general and in response to changes in the treatment variables in our experimental design.

General findings are that Bayesian inference has a substantial edge over the classical methods if one compares RMSEs of the posterior means around the data generating values with RMSEs of the classical point estimates. We find that across all model parameters in all 20 replications of all 12 experiments, the RMSEs of the SML point estimates exceed those of the MSM estimates by 21 percent on average, while the RMSEs of the MSM estimates exceed those of the posterior means obtained via Bayesian inference by 18 percent on average. Also, there is generally close agreement between the RMSEs and asymptotic standard errors for MSM, and between the RMSEs and the posterior standard deviations for Bayesian inference, but the RMSEs of the SML point estimates exceed the asymptotic standard errors by 38 percent on average. The main weakness of SML relative to the other methods is in estimation of serial correlation parameters. We examined models with AR(1) errors, and 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.

Four important patterns emerge in the response of root mean square errors 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 autocorrelation coefficients of random utility, where RMSEs fall. Second, RMSEs of Bayesian estimates relative to the classical

estimates fall significantly as serial correlation in random utility increases. Third, increasing serial correlation in the covariates leads to a small relative decrease in the RMSEs of the SML estimators. Fourth, increasing cross correlation of the errors leads to relative declines in the RMSEs of the SML and MSM estimators.

References

Allenby, G., and Rossi, P. 1993. A Bayesian approach to estimating household parameters. *Journal of Marketing Research* 30: 171–82.

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

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

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

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

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

Geweke, J., Keane, M., and Runkle, D. 1994. Alternative computational approaches to statistical inference in the multinomial probit model. Research Department Staff Report 170. Federal Reserve Bank of Minneapolis.

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

Hajivassiliou, V., and McFadden, D. 1994. A method of simulated scores for the estimation of LDV models with an application to external debt crises. Discussion Paper #967. Cowles Foundation.

Hajivassiliou, V.; McFadden, D.; and Ruud, P. 1993. Simulation of multivariate normal orthant probabilities: Methods and programs. Cowles Foundation Discussion Paper 1021. Yale University.

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

_____. 1994a. A computationally practical simulation estimator for panel data. *Econometrica* 62 (1): 95–116.

_____. 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.

_____. 1993. Asymptotic bias in maximum simulated likelihood estimation of discrete choice models. Department of Economics Working Paper #93–3. University of Michigan.

Marriott, J.; Ravishanker, N.; Gelfand, A.; and Pai, J. Forthcoming. Bayesian analysis of ARMA processes: Complete sampling based inference under exact likelihood. In *Bayesian Statistics and Applications*, ed. D. Berry, J. Geweke, and K. Chaloner. New York: Wiley.

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

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

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

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

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

		Bayesian Inference			*MSM-GHK*			*SML-GHK*		
θ	DGP	$\bar{\hat{\theta}}$	RMSE	$\overline{\text{PSD}}$	$\bar{\hat{\theta}}$	RMSE	$\overline{\text{ASE}}$	$\bar{\hat{\theta}}$	RMSE	$\overline{\text{ASE}}$
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: $\text{Corr}(\eta_{1t}, \eta_{2t}) = 0.5$, $\rho_1 = \rho_2 = 0.5$, $\text{Corr}(x_t, x_{t-1}) = 0.5$

		Bayesian Inference			*MSM-GHK*			*SML-GHK*		
θ	DGP	$\bar{\hat{\theta}}$	RMSE	$\overline{\text{PSD}}$	$\bar{\hat{\theta}}$	RMSE	$\overline{\text{ASE}}$	$\bar{\hat{\theta}}$	RMSE	$\overline{\text{ASE}}$
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

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

		Bayesian Inference			*MSM-GHK*			*SML-GHK*		
θ	DGP	$\bar{\hat{\theta}}$	RMSE	$\overline{\text{PSD}}$	$\bar{\hat{\theta}}$	RMSE	$\overline{\text{ASE}}$	$\bar{\hat{\theta}}$	RMSE	$\overline{\text{ASE}}$
a_{12}^*	0.500	0.438	0.113	0.096	0.535	0.137	0.137	0.574	0.122	0.085
a_{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

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

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

		Bayesian Inference			*MSM-GHK*			*SML-GHK*		
θ	DGP	$\overline{\hat{\theta}}$	RMSE	$\overline{\text{PSD}}$	$\overline{\hat{\theta}}$	RMSE	$\overline{\text{ASE}}$	$\overline{\hat{\theta}}$	RMSE	$\overline{\text{ASE}}$
a_{12}^*	0.500	0.506	0.053	0.079	0.567	0.121	0.107	0.620	0.138	0.067
a_{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: $\text{Corr}(\eta_{1t}, \eta_{2t}) = 0.5$, $\rho_1 = \rho_2 = 0.8$, $\text{Corr}(x_t, x_{t-1}) = 0.5$

		Bayesian Inference			*MSM-GHK*			*SML-GHK*		
θ	DGP	$\overline{\hat{\theta}}$	RMSE	$\overline{\text{PSD}}$	$\overline{\hat{\theta}}$	RMSE	$\overline{\text{ASE}}$	$\overline{\hat{\theta}}$	RMSE	$\overline{\text{ASE}}$
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: $\text{Corr}(\eta_{1t}, \eta_{2t}) = 0.5$, $\rho_1 = \rho_2 = 0.8$, $\text{Corr}(x_t, x_{t-1}) = 0.8$

		Bayesian Inference			*MSM-GHK*			*SML-GHK*		
θ	DGP	$\overline{\hat{\theta}}$	RMSE	$\overline{\text{PSD}}$	$\overline{\hat{\theta}}$	RMSE	$\overline{\text{ASE}}$	$\overline{\hat{\theta}}$	RMSE	$\overline{\text{ASE}}$
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

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

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

		Bayesian Inference			*MSM-GHK*			*SML-GHK*		
θ	DGP	$\bar{\bar{\theta}}$	RMSE	$\overline{\text{PSD}}$	$\bar{\bar{\theta}}$	RMSE	$\overline{\text{ASE}}$	$\bar{\bar{\theta}}$	RMSE	$\overline{\text{ASE}}$
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: $\text{Corr}(\eta_{1t}, \eta_{2t}) = 0.8$, $\rho_1 = \rho_2 = 0.5$, $\text{Corr}(x_t, x_{t-1}) = 0.5$

		Bayesian Inference			*MSM-GHK*			*SML-GHK*		
θ	DGP	$\bar{\bar{\theta}}$	RMSE	$\overline{\text{PSD}}$	$\bar{\bar{\theta}}$	RMSE	$\overline{\text{ASE}}$	$\bar{\bar{\theta}}$	RMSE	$\overline{\text{ASE}}$
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: $\text{Corr}(\eta_{1t}, \eta_{2t}) = 0.8$, $\rho_1 = \rho_2 = 0.5$, $\text{Corr}(x_t, x_{t-1}) = 0.8$

		Bayesian Inference			*MSM-GHK*			*SML-GHK*		
θ	DGP	$\bar{\bar{\theta}}$	RMSE	$\overline{\text{PSD}}$	$\bar{\bar{\theta}}$	RMSE	$\overline{\text{ASE}}$	$\bar{\bar{\theta}}$	RMSE	$\overline{\text{ASE}}$
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

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

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

		Bayesian Inference			*MSM-GHK*			*SML-GHK*		
θ	DGP	$\bar{\hat{\theta}}$	RMSE	PSD	$\bar{\hat{\theta}}$	RMSE	ASE	$\bar{\hat{\theta}}$	RMSE	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: $\text{Corr}(\eta_{1t}, \eta_{2t}) = 0.8$, $\rho_1 = \rho_2 = 0.8$, $\text{Corr}(x_t, x_{t-1}) = 0.5$

		Bayesian Inference			*MSM-GHK*			*SML-GHK*		
θ	DGP	$\bar{\hat{\theta}}$	RMSE	PSD	$\bar{\hat{\theta}}$	RMSE	ASE	$\bar{\hat{\theta}}$	RMSE	ASE
a_{12}^*	0.800	0.705	0.109	0.088	0.802	0.135	0.143	0.826	0.101	0.067
a_{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: $\text{Corr}(\eta_{1t}, \eta_{2t}) = 0.8$, $\rho_1 = \rho_2 = 0.8$, $\text{Corr}(x_t, x_{t-1}) = 0.8$

		Bayesian Inference			*MSM-GHK*			*SML-GHK*		
θ	DGP	$\bar{\hat{\theta}}$	RMSE	PSD	$\bar{\hat{\theta}}$	RMSE	ASE	$\bar{\hat{\theta}}$	RMSE	ASE
a_{12}^*	0.800	0.669	0.143	0.093	0.860	0.188	0.204	0.770	0.108	0.070
a_{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

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

Table 13: Comparison of Means

	Bayesian Inference			*MSM-GHK*			*SML-GHK*		
θ	BIAS	RMSE	PSD	BIAS	RMSE	ASE	BIAS	RMSE	ASE
a_{12}^*	-0.061	0.095	0.088	0.027	0.128	0.133	0.049	0.112	0.073
a_{22}^*	0.036	0.073	0.079	-0.012	0.084	0.079	0.073	0.101	0.060
ρ_1^*	-0.004	0.020	0.025	0.005	0.023	0.023	-0.035	0.042	0.019
ρ_2^*	-0.020	0.037	0.045	-0.007	0.043	0.045	-0.082	0.091	0.034
β_{11}^*	-0.007	0.023	0.037	-0.001	0.027	0.037	0.001	0.026	0.031
β_{21}^*	-0.011	0.089	0.107	0.032	0.113	0.106	-0.024	0.099	0.090
β_{12}^*	-0.018	0.036	0.043	-0.012	0.035	0.042	-0.011	0.034	0.037
β_{22}^*	-0.056	0.084	0.072	-0.019	0.082	0.080	-0.012	0.070	0.061
γ^*	-0.007	0.036	0.045	-0.021	0.044	0.043	-0.008	0.036	0.037

Note: Table reports the mean of the RMSE, the BIAS, and the PSD or ASE for each model parameter across the 12 experiments.

Table 14: 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
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
Corr $(\eta_{1t}, \eta_{2t}) = 0.80$ X DEP=RHO	0.051	0.049	1.043
$\rho_1 = \rho_2 = 0.80$ X DEP=RHO	-0.421	0.049	-8.615
$\phi^2 = 0.50$ X DEP=RHO	-0.383	0.060	-6.409
$\phi^2 = 0.80$ X DEP=RHO	-0.530	0.060	-8.864
a_{12}^* X MSM-GHK	0.103	0.105	0.980
a_{22}^* X MSM-GHK	-0.047	0.105	-0.447
ρ_1^* X MSM-GHK	-0.021	0.105	-0.196
ρ_2^* X MSM-GHK	-0.034	0.105	-0.326
β_{11}^* X MSM-GHK	-0.015	0.105	-0.140
β_{21}^* X MSM-GHK	0.061	0.105	0.574
β_{12}^* X MSM-GHK	-0.181	0.105	-1.713
β_{22}^* X MSM-GHK	-0.195	0.105	-1.852
a_{12}^* X SML-GHK	0.191	0.105	1.812
a_{22}^* X SML-GHK	0.280	0.105	2.650
ρ_1^* X SML-GHK	0.767	0.105	7.274
ρ_2^* X SML-GHK	0.909	0.105	8.621
β_{11}^* X SML-GHK	0.110	0.105	1.043
β_{21}^* X SML-GHK	0.108	0.105	1.028
β_{12}^* X SML-GHK	-0.046	0.105	-0.434
β_{22}^* X SML-GHK	-0.167	0.105	-1.585
Corr $(\eta_{1t}, \eta_{2t}) = 0.80$ X MSM-GHK	-0.082	0.050	-1.652
$\rho_1 = \rho_2 = 0.80$ X MSM-GHK	0.129	0.050	2.597
$\phi^2 = 0.50$ X MSM-GHK	0.033	0.061	0.543
$\phi^2 = 0.80$ X MSM-GHK	-0.057	0.061	-0.938
Corr $(\eta_{1t}, \eta_{2t}) = 0.80$ X SML-GHK	-0.129	0.050	-2.585
$\rho_1 = \rho_2 = 0.80$ X SML-GHK	0.142	0.050	2.865
$\phi^2 = 0.50$ X SML-GHK	-0.018	0.061	-0.292
$\phi^2 = 0.80$ X 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.

$$\epsilon_{1,t} = \rho_1 \epsilon_{1,t-1} + \eta_1, \qquad \epsilon_{2,t} = \rho_2 \epsilon_{2,t-1} + \eta_2, \qquad \epsilon_{3,t} = 0$$