



WP/09/74

IMF Working Paper

Limited Information Bayesian Model Averaging for Dynamic Panels with Short Time Periods

Huigang Chen, Alin Mirestean, and Charalambos Tsangarides

IMF Working Paper

Research Department

**Limited Information Bayesian Model Averaging for Dynamic Panels
with Short Time Periods**

Prepared by Huigang Chen, Alin Mirestean, and Charalambos Tsangarides ¹

Authorized for distribution by Atish Ghosh

April 2009

Abstract

This Working Paper should not be reported as representing the views of the IMF.

The views expressed in this Working Paper are those of the author and do not necessarily represent those of the IMF or IMF policy. Working Papers describe research in progress by the authors and are published to elicit comments and to further debate.

Bayesian Model Averaging (BMA) provides a coherent mechanism to address the problem of model uncertainty. In this paper we extend the BMA framework to panel data models where the lagged dependent variable as well as endogenous variables appear as regressors. We propose a Limited Information Bayesian Model Averaging (LIBMA) methodology and then test it using simulated data. Simulation results suggest that asymptotically our methodology performs well both in Bayesian model selection and averaging. In particular, LIBMA recovers the data generating process very well, with high posterior inclusion probabilities for all the relevant regressors, and parameter estimates very close to the true values. These findings suggest that our methodology is well suited for inference in dynamic panel data models with short time periods in the presence of endogenous regressors under model uncertainty.

JEL Classification Numbers: C11, C15, C23, C52.

Keywords: Bayesian Model Averaging; Model Uncertainty; Dynamic Panels; Generalized Method of Moments; Robustness.

Author's E-Mail Address: hchen@imf.org; amirestean@imf.org; ctsangarides@imf.org.

¹ We thank Steve Durlauf, Rex Ghosh, Eduardo Ley, Enrique Moral-Benito, Chris Papageorgiou, and Roberto Samaniego for helpful comments and suggestions.

	Page
I. Introduction	2
II. Model Uncertainty in the Bayesian Context	3
A. Model Selection and Hypothesis Testing	3
B. Bayesian Model Averaging	4
C. Choice of Priors	5
III. Limited Information Bayesian Model Averaging	7
A. A Dynamic Panel Data Model with Endogenous Regressors	7
B. Estimation and Moment Conditions	8
C. The Limited Information Criterion	12
IV. Monte Carlo Simualtions and Results	16
A. The Data Generating Process	16
B. Simulation Results	17
V. Conclusion	21
References	22
Tables:	
1. Posterior Probability of the True Model	27
2. Posterior Probability Ratio of True Model/Best among the Other Models	28
3. Probability of Retrieving the True Model	29
4. Model Recovery: Medians and Variances of Posterior Inclusion Probability for Each Variable	30
5. Model Recovery: Medians and Variances of Estimated Parameter Values	31
6. Posterior Probability of the True Model (Non-Gaussian Case)	32
7. Posterior Probability Ratio: True Model/best among the Other Models (Non-Gaussian Case)	33
8. Probability of Retrieving the True Model (Non-Gaussian Case).....	34
9. Model Recovery: Medians and Variances of Posterior Inclusion Probability for Each Variable (Non-Gaussian Case)	35
10. Model Recovery: Medians and Variances of Estimated Parameter Values (Non-Gaussian Case)	36
Appendix A Figures:	
1. Posterior Densities for the Probabilities in Table 1	37
2. Posterior Densities for the Probabilities in Table 2	38
3. Box Plots for Parameters in Table 5	39
4. Posterior Densities for the Probabilities in Table 6	40
5. Posterior Densities for the Probabilities in Table 7	41
6. Box Plots for Parameters in Table 10	42

I. INTRODUCTION

Model uncertainty is an issue encountered often in the econometric study of socioeconomic phenomena. Initially pointed out by Leamer (1978) and later elaborated by Durlauf and Quah (1999) model uncertainty arises because the lack of clear theoretical guidance and tradeoffs on the choice of regressors result in a broad number of possible specifications, and often contradictory conclusions. In addition, attempts to deal with model uncertainty by engaging in unsystematic searches of possible model configurations, may result in overconfident and often fragile inferences. As a result, a growing number of researchers are turning to the Bayesian Model Averaging (BMA) framework in order to deal with the problem of model uncertainty.

Conceptually, BMA bases inferences on a weighted average of the full model space instead of on one selected model, and thus incorporates uncertainty in both predictions and parameter estimates.² Seminal contributions to BMA include those of Moulton (1991), Madigan and Raftery (1994), Kass and Raftery (1995), Raftery (1995), and Raftery, Madigan and Hoeting (1997). The BMA framework has been applied in various areas of social sciences.³ In economics, some of the most notable work includes Brock and Durlauf (2001), Fernández, Ley and Steel (2001a), and Sala-i-Martin, Doppelhofer and Miller (2004). Despite the increasing interest in BMA, most of the work thus far uses static models, focusing mainly on cross section analysis with data averaged over the time dimension, thus ignoring dynamic relationships among variables.⁴ Moreover, to the best of our knowledge, none of the models allow for the inclusion of endogenous variables.⁵

In this paper we propose a methodology for dealing with model uncertainty in the context of panel data model with short time periods where the lagged dependent variable and endogenous variables appear as regressors. We use a limited information approach which refines the limited information version of Bayesian Model Averaging (LIBMA) introduced by Tsangarides (2004). The limited information criterion proposed in this paper resembles the

² In contrast to BMA, Bayesian model selection uses information criteria to select one model (one set of variables) from a set of potential models.

³ These include biology (Yeung, Bungarner, and Raftery (2005)), ecology (Wintle et al. (2003)), public health (Morales et al. (2006)), and toxicology (Koop and Tole (2004)).

⁴ Moral-Benito (2007) considers a panel data model where the lagged dependent variable is correlated with the individual effects but not correlated with the error term.

⁵ By endogenous variables we understand variables that would be correlated with the contemporary error term.

BIC model and moment selection criterion (MMSC-BIC) proposed by Andrews and Lu (2001), and parallels the one proposed by Hong and Preston (2008). One key difference of our approach is that we construct the likelihood by data transformation and straightforward Bayesian arguments.⁶ We also investigate the performance of the proposed framework relative to both Bayesian model selection and averaging by performing Monte Carlo simulations.

The remainder of the paper is organized as follows. In Section 2 we introduce the concept of model uncertainty in the Bayesian context and then review model selection and model averaging. Section 3 develops the theoretical framework of the LIBMA methodology in the context of dynamic panels with endogenous regressors. It includes the model setup, the moment conditions, the limited information criterion, and estimation. Section 4 discusses the proposed simulation experiment and presents the results. The final section concludes.

II. MODEL UNCERTAINTY IN THE BAYESIAN CONTEXT

For completeness, this section reviews briefly the basic theory of uncertainty in the Bayesian context. Excellent reviews include Hoeting, Madigan, Raftery and Volinsky (1999), and Chipman, George and McCulloch (2001).

A. Model Selection and Hypothesis Testing

Consider the standard linear regression model

$$Y = \tilde{Z}\theta + u \tag{1}$$

where Y is the variable of interest, \tilde{Z} is a matrix of explanatory variables, θ is a vector of unknown parameters and u is the error term. Suppose there is a universe of k possible explanatory variables indexed by $\mathbf{U} = \{1, 2, \dots, j, j+1, \dots, k\}$. Let \mathbf{Z} be the matrix of all possible explanatory variables. For a given model M_j that considers only a subset of the possible explanatory variables, $M_j \subset \mathbf{U}$, let $\mathbf{C}_{M_j} = \{c_{mn, M_j}\}_{m,n=1}^k$ be a $k \times k$ diagonal

⁶ Another related approach to derive Bayesian posteriors in models with moment conditions is by using information projection onto a family of probability measures. Kim (2002) uses a transformation of the GMM objective function to obtain a likelihood that is then used for deriving the posterior. Ragusa (2008) uses the projection of a reference distribution into the space of distributions that are consistent with the average form of the moment conditions and then uses formal Bayesian calculus to obtain the posterior.

choice matrix such that its diagonal will have 1's if the corresponding variable is included in the model and 0's otherwise. Hence $c_{ii,M_j} = 1\{i \in M_j\}$. Therefore, for a given model M_j , $\tilde{\mathbf{Z}} = \mathbf{Z}\mathbf{C}_{M_j}$ and model (1) can be now written more generally as

$$Y = \mathbf{Z}\mathbf{C}_{M_j}\theta + u \quad (2)$$

where $\theta = (\theta_1 \ \theta_2 \ \dots \ \theta_k)'$ is the set of parameters to be estimated.

Given the universe of k possible explanatory variables, a set of $K = 2^k$ models $\mathcal{M} = (M_1, \dots, M_K)$ are under consideration. In the spirit of Bayesian inference, one can specify priors $p(\theta \mid M_j)$ for the parameters of each model, and a prior $p(M_j)$ for each model in the model space \mathcal{M} .

Model selection seeks to find the model M_j in $\mathcal{M} = (M_1, \dots, M_K)$ that actually generated the data. Let $D = (\mathbf{Y} \ \mathbf{Z})$ denote the data set available to the researcher. The probability that M_j is the correct model, given the data D , is, by Bayes' rule

$$p(M_j \mid D) = \frac{p(D \mid M_j)p(M_j)}{\sum_{l=1}^K p(D \mid M_l)p(M_l)} \quad (3)$$

where

$$p(D \mid M_j) = \int p(D \mid \theta_j, M_j)p(\theta_j \mid M_j)d\theta_j \quad (4)$$

is the marginal probability of the data given model M_j .

Based on the posterior probabilities, the comparison of model M_j against M_i is expressed by the posterior odds ratio $\frac{p(M_j \mid D)}{p(M_i \mid D)} = \frac{p(D \mid M_j)}{p(D \mid M_i)} \cdot \frac{p(M_j)}{p(M_i)}$. Essentially, the data updates the prior odds ratio $\frac{p(M_j)}{p(M_i)}$ through the Bayes factor $\frac{p(D \mid M_j)}{p(D \mid M_i)}$ to measure the extent to which the data support M_j over M_i . When the posterior odds ratio is greater (less) than 1 the data favor M_j over M_i (M_i over M_j). Often the prior odds ratio is set to

1 representing the lack of preference for either model,⁷ in which case the posterior odds ratio is equal to the Bayes factor B_{ji} .

B. Bayesian Model Averaging

A natural strategy for model selection is to choose the most probable model M_j , namely the one with the highest posterior probability, $p(M_j | D)$. Alternatively, especially in cases where the posterior mass of the model space \mathcal{M} is not concentrated only on one model, M_j , it is possible to consider averaging models using the posterior model probabilities as weights. Raftery, Madigan, and Hoeting (1997) show that BMA almost always improves on variable selection in terms of predictive performance.

Using Bayesian Model Averaging, inference for a quantity of interest Γ can be constructed based on the posterior distribution

$$p(\Gamma | D) = \sum_{j=1}^K p(\Gamma | D, M_j) p(M_j | D) \quad (5)$$

which follows by the law of total probability. Therefore, the full posterior distribution of Γ is a weighted average of the posterior distributions under each model (M_1, \dots, M_K) , where the weights are the posterior model probabilities $p(M_j | D)$. The posterior model probabilities are obtained using (3). Using (5) one can compute the posterior mean and posterior variance for parameters θ_l as follows

$$E(\theta_l | D) = \sum_{j=1}^K p(M_j | D) E(\theta_l | D, M_j) \quad (6)$$

and

$$Var(\theta_l | D) = \sum_{j=1}^K p(M_j | D) Var(\theta_l | D, M_j) + \sum_{j=1}^K p(M_j | D) [E(\theta_l | D, M_j) - E(\theta_l | D)]^2. \quad (7)$$

The implementation of BMA presents a number of challenges, including the evaluation of the marginal probability in (4), the large number of possible models, and the specification of the prior model probabilities $p(M_j)$ as well as the parameters' prior, $p(\theta | M_i)$.

⁷ As in Fernández, Ley and Steel (2001b).

C. Choice of Priors

Evaluating Bayes factors required for hypothesis testing and Bayesian model selection or model averaging requires calculating the marginal likelihood

$$p(D | M_j) = \int p(D | \theta, M_j) p(\theta | M_j) d\theta. \quad (8)$$

Here, the dimension of the parameter θ is determined by model M_j . In many cases the likelihood $p(D | \theta, M_i)$ is fully specified with some nuisance parameter ζ . Therefore we may write

$$p(D | M_i) = \int p(D | \theta, \zeta, M_i) p(\theta, \zeta | M_i) d\theta d\zeta. \quad (9)$$

In this case, determining the prior $p(\theta, \zeta | M_i)$ becomes an important issue.⁸

For Gaussian models the nuisance parameter is the variance σ_u^2 of the noise term. A common selection of the prior for the pair (θ, σ_u^{-2}) is its conjugate prior, Normal-Gamma distribution, which has the benefit of rendering a closed-form posterior⁹. With this prior θ is a Normal random variable with mean θ_0 and variance $\sigma_u^2 V$ given σ_u^2 , while σ_u^{-2} is a Gamma random variable with mean $\frac{\gamma}{\lambda}$ and variance $\frac{\gamma}{\lambda^2}$. Based on this prior, when (1) represents a Gaussian panel data model with fixed effects, the likelihood ratio of two different models, M_i and M_j , becomes

$$\begin{aligned} \frac{p(D|M_i)}{p(D|M_j)} &= \left(\frac{|I + \tilde{Z}_j' \tilde{Z}_j V_j|}{|I + \tilde{Z}_i' \tilde{Z}_i V_i|} \right)^{1/2} \\ &\times \left(\frac{2\lambda + SSE_j + (\hat{\theta}_j - \theta_j)' (V_j + (\tilde{Z}_j' \tilde{Z}_j)^{-1})^{-1} (\hat{\theta}_j - \theta_j)}{2\lambda + SSE_i + (\hat{\theta}_i - \theta_i)' (V_i + (\tilde{Z}_i' \tilde{Z}_i)^{-1})^{-1} (\hat{\theta}_i - \theta_i)} \right)^{(\gamma + N(T-1))/2} \end{aligned}$$

where \tilde{Z} stands for the demeaned values of Z . Due to the sensitivity of the Bayes factors to the prior parameters $\{\theta_0, V, \gamma, \lambda\}$, one often avoids choosing specific values for them, in order not to affect substantially the posterior distribution. As discussed in Kass and Wasserman (1995), and Fernández, Ley and Steel (2001a), one possibility is to use a diffuse

⁸ Fernández, Ley and Steel (2001b) investigate a set of “benchmark” prior specifications in a linear regression context with model uncertainty in order to address the sensitivity of the posterior model probabilities to the specification of the priors.

⁹ For a more detailed discussion see Kass and Wasserman (1995).

prior for σ_u with density $p(\sigma_u) \propto \sigma_u^{-1}$. This prior has a nice scale invariance property and is equivalent to setting $\gamma = \lambda = 0$ in the Gamma distribution of σ_u^{-2} . For the prior distribution of θ conditioned on σ_u^{-2} , one popular choice is using Zellner's g -prior with 0 mean

$$p(\theta \mid \sigma_u^2) \sim N\left(0, g^{-1}(\tilde{Z}'\tilde{Z})^{-1} \sigma_u^2\right)$$

which can be motivated by the fact that the correlation of the OLS estimate $\hat{\theta}$ is proportional to $(\tilde{Z}'\tilde{Z})^{-1} \sigma_u^2$. This also leads to a simple likelihood ratio

$$\frac{p(D \mid M_i)}{p(D \mid M_j)} = \frac{(1 + g_i^{-1})^{k_i/2} \left(\frac{1}{g_j+1} SSE_j + \frac{g_j}{g_j+1} \tilde{y}'\tilde{y} \right)^{N(T-1)/2}}{(1 + g_j^{-1})^{k_j/2} \left(\frac{1}{g_i+1} SSE_i + \frac{g_i}{g_i+1} \tilde{y}'\tilde{y} \right)^{N(T-1)/2}}$$

where \tilde{y} stands for the demeaned values of y . The BACE procedure (proposed by Sala-i-Martin, Doppelhofer, and Miller (2004)) is asymptotically equivalent to setting $g = N(T-1)$.

Alternatively, one can use what has been labeled in the literature as the BIC approach, where the likelihood ratio is approximated by

$$\frac{p(D \mid M_1)}{p(D \mid M_2)} \sim \exp\left(-\frac{1}{2}(SSE_1 - SSE_2) - \frac{1}{2}(k_1 - k_2) \log(N(T-1))\right).$$

Here the approximation is $O_p(N^{-1/2})$ when the implicit prior for θ is the unit information Normal prior as discussed in Kass and Wasserman (1995) and Kass and Raftery (1995).

Finally, several options exist for the specification of the model priors $p(M_j)$. For example, Fernández, Ley and Steel (2001b) assume a Uniform distribution over the model space, essentially implying that there is no preference for a specific model so $p(M_1) = p(M_2) = \dots = p(M_K) = \frac{1}{K}$. Other options include penalizing models with more regressors. Sala-i-Martin, Doppelhofer and Miller (2004) use a prior model probability structure initially proposed by Mitchell and Beauchamp (1988). Assuming that each variable has an equal inclusion probability, the prior probability for model M_j is

$$p(M_j) = \left(\frac{\bar{k}}{k}\right)^{k_j} \left(1 - \frac{\bar{k}}{k}\right)^{k-k_j} \quad (10)$$

and the prior odds ratio is

$$\frac{p(M_j)}{p(M_l)} = \left(\frac{\bar{k}}{k}\right)^{k_j - k_l} \left(1 - \frac{\bar{k}}{k}\right)^{k_l - k_j} \quad (11)$$

where k is the total number of regressors, \bar{k} is the researcher's prior about the size of the model, k_j is the number of included variables in model M_j , and $\frac{\bar{k}}{k}$ is the prior inclusion probability for each variable.

III. LIMITED INFORMATION BAYESIAN MODEL AVERAGING

This section provides a discussion of the LIBMA using a dynamic panel data model with endogenous and exogenous regressors and derives the limited information criterion using the moment conditions implied by the GMM framework.

A. A Dynamic Panel Data Model with Endogenous Regressors

Let us consider the case where a researcher is faced with model uncertainty when trying to estimate a dynamic model for panel data. We assume that the universe of potential explanatory variables, indexed by the set \mathbf{U} , consists of the lagged dependent variable, indexed by 1, a set of m exogenous variables, indexed by \mathbf{X} , as well as a set of q endogenous variables, indexed by \mathbf{W} , such that $\{\{1\}, \mathbf{X}, \mathbf{W}\}$ is a partition of \mathbf{U} . Therefore, for a given model $M_j \subset \mathbf{U}$, (2) becomes

$$\begin{aligned} y_{it} &= \begin{pmatrix} y_{i,t-1} & \mathbf{x}_{it} & \mathbf{w}_{it} \end{pmatrix} \mathbf{C}_{M_j} \begin{pmatrix} \alpha & \theta_x & \theta_w \end{pmatrix}' + u_{it} \\ u_{it} &= \eta_i + v_{it} \\ |\alpha| &< 1; \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T. \end{aligned} \quad (12)$$

Here y_{it} , \mathbf{x}_{it} and \mathbf{w}_{it} are observed variables, η_i is the unobserved individual effect while v_{it} is the idiosyncratic random error. The exact distributions for v_{it} and η_i are not specified here, but assumptions about some of their moments and correlation with the regressors are made explicit below. It is assumed that $E(v_{it}) = 0$ and that v_{it} 's are not serially correlated. \mathbf{x}_{it} is a $1 \times m$ vector of exogenous variables while \mathbf{w}_{it} is a $1 \times q$ vector of endogenous variables. Therefore the total number of possible explanatory variables is $k = m + q + 1$. The observed variables span N individuals and T periods, where T is small relative to N . The unknown parameters α , θ_x , and θ_w are to be estimated. In this model, α is a scalar, θ_x is a $1 \times m$ vector while θ_w is a $1 \times q$ vector.

Given the assumptions made so far, for any model M_j , and any set of exogenous variables, \mathbf{x}_{it} , we have

$$E(x_{it}^l v_{is}) = 0, \quad \forall i, t, s; \quad x_{it}^l \in \mathbf{x}_{it}. \quad (13)$$

Similarly, for any endogenous variable we have

$$E(w_{it}^l v_{is}) \begin{cases} \neq 0, & s \leq t \\ = 0, & \text{otherwise} \end{cases}, \quad w_{it}^l \in \mathbf{w}_{it}. \quad (14)$$

Note that, in principle, the correlations between endogenous variables and the idiosyncratic error may change over different individuals and/or periods.

B. Estimation and Moment Conditions

A common approach for estimating the model (12) is to use the system GMM framework developed by Blundell and Bond (1998). This implies constructing the instruments set and moment conditions for the “level equations” (12) and combining them with the moment conditions using the instruments corresponding to the first differences equations. The first differences (FD) equations corresponding to model (12) are given by

$$\begin{aligned} \Delta y_{it} &= \begin{pmatrix} \Delta y_{i,t-1} & \Delta \mathbf{x}_{it} & \Delta \mathbf{w}_{it} \end{pmatrix} \mathbf{C}_{M_j} \begin{pmatrix} \alpha & \theta_x & \theta_w \end{pmatrix}' + \Delta v_{it} \\ |\alpha| &< 1; \quad i = 1, 2, \dots, N; \quad t = 2, 3, \dots, T. \end{aligned} \quad (15)$$

One assumption required for the FD equations is that the initial value of \mathbf{y} , y_{i0} , is predetermined, that is, $E(y_{i0} v_{is}) = 0$ for $s = 2, 3, \dots, T$. Since $y_{i,t-2}$ is not correlated with Δv_{it} we can use it as an instrument. Hence we have $E(y_{i,t-2} \Delta v_{it}) \neq 0$ for $t = 2, 3, \dots, T$. Moreover, $y_{i,t-3}$ is also not correlated with Δv_{it} . Therefore, as long as we have enough observations, that is $T \geq 3$, $y_{i,t-3}$ can be used as an instrument. Assuming that we have more than two observations in the time dimension, the following moment conditions could be used for estimation

$$E(y_{i,t-s} \Delta v_{it}) = 0, \quad t = 2, 3, \dots, T; \quad s = 2, 3, \dots, t; \quad \text{for } T \geq 2, \quad i = 1, 2, \dots, N. \quad (16)$$

Similarly, the first difference of the exogenous variable Δx_{it}^l , $x_{it}^l \in \mathbf{x}_{it}$ is not correlated with Δv_{it} and therefore we can use it as an instrument.¹⁰ That gives us additional moment conditions

$$E(\Delta x_{it}^l \Delta v_{it}) = 0, \quad t = 2, 3, \dots, T; \quad l = 1, \dots, m; \quad i = 1, 2, \dots, N. \quad (17)$$

The endogenous variable $w_{i,t-2}^l$, $w_{i,t-2}^l \in \mathbf{w}_{it}$, is not correlated with Δv_{it} and therefore it can be used as an instrument. We have the following possible moment conditions

$$\begin{aligned} E(w_{i,t-s}^l \Delta v_{it}) &= 0, \quad t = 3, 4, \dots, T; \quad s = 2, \dots, t-1; \\ &\text{for } T \geq 3, \quad l = 1, 2, \dots, q; \quad i = 1, \dots, N. \end{aligned} \quad (18)$$

Table 1 summarizes the moment conditions that could be used for the FD equation.

Table 1. Moment Conditions for the FD Equation

	Instruments	Moment conditions
Lagged dependent variable $\Delta y_{i,t-1}$	$y_{i,t-2}, \dots, y_{i0}$	$E(y_{i,t-s} \Delta v_{it}) = 0, \quad t = 2, 3, \dots, T; \quad s = 2, 3, \dots, t$
Exogenous variable Δx_{it}^l	$x_{it}^l, \dots, x_{i1}^l$	$E(\Delta x_{it}^l \Delta v_{it}) = 0, \quad t = 2, 3, \dots, T; \quad l = 1, \dots, m$
Endogenous variable Δw_{it}^l	$w_{i,t-2}^l, \dots, w_{i,1}^l$	$E(w_{i,t-s}^l \Delta v_{it}) = 0, \quad t = 3, 4, \dots, T; \\ s = 2, 3, \dots, t-1; \quad l = 1, 2, \dots, q$

The FD equation provides $T(T-1)/2$ moment conditions for the lagged dependent variable, $m(T-1)$ moment conditions for the exogenous variables, and $q(T-2)(T-1)/2$ moment conditions for the endogenous variables.

Going back to the equation in levels (12), it is easy to see that first differences for the lagged dependent variable are not correlated with either the individual effects or the idiosyncratic error term and hence we can use the following moment conditions

$$E(\Delta y_{i,t-1} u_{it}) = 0, \quad t = 2, 3, \dots, T. \quad (19)$$

¹⁰ It is common in the literature to use $x_{it}^l \in \mathbf{x}_{it}$ as an instrument, instead of Δx_{it}^l . Then the moment condition (17) becomes $E(x_{it}^l \Delta v_{it}) = 0$.

Similarly, for the endogenous variables the first difference $\Delta w_{i,t-1}^l$ is not correlated with u_{it} . Therefore, assuming that $w_{i,0}^l$ is observable, and as long as $T \geq 3$ we have the following additional moment conditions

$$E(\Delta w_{i,t-1}^l u_{it}) = 0, \quad t = 3, 4, \dots, T; \quad l = 1, 2, \dots, q. \quad (20)$$

Finally, based on the assumptions made so far, the exogenous variables $x_{it}^l \in \mathbf{x}_{it}$ are not correlated with current realizations of u_{it} and hence one can use another set of moment conditions¹¹

$$E(x_{it}^l u_{it}) = 0, \quad t = 1, 2, \dots, T; \quad l = 1, 2, \dots, m. \quad (21)$$

Table 2 summarizes the moment conditions for the level equation.

Table 2. Moment Conditions for the Level Equation

	Instruments	Moments
Lagged dependent variable $y_{i,t-1}$	$\Delta y_{i,t-1}$	$E(\Delta y_{i,t-1} u_{it}) = 0, \quad t = 2, 3, \dots, T$
Exogenous variable x_{it}^l	x_{it}^l	$E(x_{it}^l u_{it}) = 0, \quad t = 1, 2, \dots, T; \quad l = 1, 2, \dots, m$
Endogenous variable w_{it}^l	$\Delta w_{i,t-1}^l$	$E(\Delta w_{i,t-1}^l u_{it}) = 0, \quad t = 3, 4, \dots, T; \quad l = 1, 2, \dots, q$

The equation in levels provides $(T - 1)$ moment conditions for the lagged dependent variable, mT moment conditions for the exogenous variables, and $q(T - 2)$ moment conditions for the endogenous variables.

We group the moment conditions into matrices the following way. Let Y_i be the $(T - 1) \times T(T - 1)/2$ matrix of lagged dependent variable used as instruments for the FD equation

¹¹ The implicit assumption made here is that the exogenous variables are not correlated with the individual effect. If that assumption can not be made, Δx_{it}^l should be used as instruments, giving us the following moment conditions $E(\Delta x_{it}^l u_{it}) = 0, \quad t = 1, 2, \dots, T; \quad l = 1, 2, \dots, m$.

$$Y_i = \begin{pmatrix} y_{i0} & 0 & 0 & 0 & 0 & 0 & \cdots 0 \\ 0 & y_{i0} & y_{i1} & 0 & 0 & 0 & \cdots 0 \\ 0 & 0 & 0 & y_{i0} & y_{i1} & y_{i2} & \cdots 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots y_{i0} \cdots y_{i,T_i-2} \end{pmatrix}. \quad (22)$$

Similarly, W_i denotes the $(T-1) \times q(T-2)(T-1)/2$ matrix of endogenous variables

$$W_i = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & \cdots 0 & 0 \\ w_{i1}^1 & 0 & 0 & \cdots & w_{i1}^q & \cdots 0 & 0 \\ 0 & w_{i1}^1 & w_{i2}^1 & \cdots & 0 & \cdots 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \cdots 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \cdots w_{i,T-3}^q & w_{i,T-2}^q \end{pmatrix}. \quad (23)$$

For the level equation we have the $T \times (T-1)$ instruments matrix DY_i consisting of first differences of the dependent variable and the $T \times q(T-2)$ instruments matrix DW_i consisting of first differences of the endogenous variables

$$DY_i = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ \Delta y_{i1} & 0 & 0 & \cdots & 0 \\ 0 & \Delta y_{i2} & 0 & \cdots & 0 \\ 0 & 0 & \Delta y_{i3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \Delta y_{i,T-1} \end{pmatrix} \quad DW_i = \begin{pmatrix} 0 & 0 \cdots & 0 \\ 0 \cdots & 0 \cdots & 0 \\ \Delta w_{i2}^1 \cdots & \Delta w_{i2}^q \cdots & 0 \\ 0 \cdots & 0 \cdots & 0 \\ \vdots & \cdots & \vdots \\ 0 \cdots & 0 \cdots & \Delta w_{i,T-1}^q \end{pmatrix}. \quad (24)$$

Further let X_i and DX_i denote the following $T \times m$ and $(T-1) \times m$ matrices of exogenous and first differenced exogenous variables, respectively

$$DX_i = \begin{pmatrix} \Delta x_{i2}^1 & \Delta x_{i2}^2 & \Delta x_{i2}^3 & \cdots & \Delta x_{i2}^m \\ \Delta x_{i3}^1 & \Delta x_{i3}^2 & \Delta x_{i3}^3 & \cdots & \Delta x_{i3}^m \\ \Delta x_{i4}^1 & \Delta x_{i4}^2 & \Delta x_{i4}^3 & \cdots & \Delta x_{i4}^m \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \Delta x_{iT}^1 & \Delta x_{iT}^2 & \Delta x_{iT}^3 & \cdots & \Delta x_{iT}^m \end{pmatrix} \quad X_i = \begin{pmatrix} x_{i1}^1 & x_{i1}^2 & x_{i1}^3 & \cdots & x_{i1}^m \\ x_{i2}^1 & x_{i2}^2 & x_{i2}^3 & \cdots & x_{i2}^m \\ x_{i3}^1 & x_{i3}^2 & x_{i3}^3 & \cdots & x_{i3}^m \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ x_{iT}^1 & x_{iT}^2 & x_{iT}^3 & \cdots & x_{iT}^m \end{pmatrix}. \quad (25)$$

For the exogenous variables, we aggregate the moment conditions across all periods from both the first difference equation and the level equation. Thus, we are left with one moment condition for each of the exogenous variables

$$\sum_{t=2}^T E(\Delta x_{it}^l \Delta v_{it}) + \sum_{t=1}^T E(x_{it}^l u_{it}) = 0, \quad l = 1, \dots, m; \quad i = 1, 2, \dots, N.$$

Let u_i and Dv_i denote the $T \times 1$ and $(T-1) \times 1$ matrices of the error term and the first differenced idiosyncratic random error, respectively, as defined in model (12).

$$u_i = \begin{pmatrix} u_{i1} & u_{i2} & \cdots & u_{iT} \end{pmatrix}' \quad Dv_i = \begin{pmatrix} \Delta v_{i2} & \Delta v_{i3} & \cdots & \Delta v_{iT} \end{pmatrix}'. \quad (26)$$

We can define a $(2T-1) \times 1$ matrix $U_i = \begin{pmatrix} u_i' & Dv_i' \end{pmatrix}'$ that contains both the error term and the first differenced idiosyncratic random error. The moment conditions can now be written in matrix form

$$E[G_i' U_i] = 0 \quad (27)$$

where G_i is a $(2T-1) \times (m-1 + (T+1)((T-2)q + T)/2)$ matrix defined as

$$G_i = \begin{pmatrix} X_i & DY_i & 0_{T \times T(T-1)/2} & DW_i & 0_{T \times q(T-1)(T-2)/2} \\ DX_i & 0_{(T-1) \times (T-1)} & Y_i & 0_{(T-1) \times q(T-2)} & W_i \end{pmatrix}. \quad (28)$$

Based on the moment conditions (27) we propose a limited information criterion that can be used in Bayesian model selection and averaging. In the next section we provide details on how to construct this criterion.

C. The Limited Information Criterion

As we pointed out in section II.C, evaluating the Bayes factors needed for hypothesis testing and Bayesian model selection or model averaging requires calculating the marginal likelihood

$$p(D | M_j) = \int p(D | \theta, M_j) p(\theta | M_j) d\theta. \quad (29)$$

Given that we choose to use the Generalized Method of Moments (GMM) for estimating the parameters of the model, the assumptions we have made so far do not give us a fully specified parametric likelihood $p(D | \theta, M_i)$. Therefore, we have to build the model likelihood in a fashion consistent with the Bayesian paradigm using the information provided by the moment conditions.

The construction of non-parametric likelihood functions has received lately a good deal of attention in the literature. Several approaches have been used to derive or estimate non-parametric likelihood functions. For example, Back and Brown (1993) provide a method of estimating a distribution function using only information derived from moment restrictions. Kim (2002) uses information projection onto a family of probability measures and constructs the likelihood by using a transformation of the GMM objective function. Hong and Preston (2008) build a quasi likelihood which is based on objective functions used for extremum estimation (see also Chernozhukov and Hong (2003)). Schennach (2005) builds a likelihood function that is the nonparametric limit result of a formal Bayesian procedure where the prior for the data favors distributions with a large entropy. Further the prior is conditioned on the moment equations. In this fashion it becomes feasible to compute a likelihood function that is closely related to empirical likelihood. Finally, Ragusa (2008) projects a reference distribution onto the space of distributions that are consistent with a set of moment restrictions and obtains the likelihood by integrating out the nuisance parameters.

In this section we propose a method of constructing the model likelihoods and posteriors based only on the information elicited from the moment conditions (27). Suppose we have a strictly stationary and ergodic random process $\{\xi_i\}_{i=1}^{\infty}$, which takes value in the space Ξ , and a parameter space $\Theta \subset R^k$. Then there exists a function $g : \Xi \times \Theta \rightarrow R^l$ which satisfies the following conditions

- (1) it is continuous on Θ ;
- (2) $E[g(\xi_i, \theta)]$ exists and is finite for every $\theta \in \Theta$; and
- (3) $E[g(\xi_i, \theta)]$ is continuous on θ .

We further assume that the moment conditions, $E[g(\xi_i, \theta)] = 0$, hold for a unique unknown

$\theta_0 \in \Theta$. Let $\hat{g}_N(\theta) = N^{-1} \sum_{i=1}^N g(\xi_i, \theta)$ denote the sample mean of the moment conditions,

and assume that $E[g(\xi_i, \theta_0)g'(\xi_i, \theta_0)]$ and $S(\theta_0) \equiv \lim_{n \rightarrow \infty} Var[N^{1/2}\hat{g}_N(\theta_0)]$ exist and are finite positive definite matrices. Then the following standard result holds (for a proof see Hall (2005) Lemma 3.2).

Lemma 1. Under the above assumptions, $N^{1/2}\hat{g}_N(\theta_0) \xrightarrow{d} N(0, S(\theta_0))$.

That is, the random vector $N^{1/2}\hat{g}_N(\theta_0)$ converges in distribution to a multivariate Normal distribution.

For model (12), the moment conditions for individual i discussed in the previous section can be written in the following form

$$g(\xi_i, \theta) = G_i' (\tilde{\mathbf{y}}_i - \tilde{\mathbf{z}}_i \theta) \quad (30)$$

where $\xi_i = \{\tilde{\mathbf{y}}_i, \tilde{\mathbf{z}}_i\}$, $\tilde{\mathbf{z}}_i = (\tilde{\mathbf{y}}_{i,-1} \quad \tilde{\mathbf{x}}_i \quad \tilde{\mathbf{w}}_i)$, $\theta = (\alpha \quad \theta_x \quad \theta_w)'$, while G_i is the matrix defined in (28). The vectors $\tilde{\mathbf{y}}_i$ and $\tilde{\mathbf{y}}_{i,-1}$ for the dependent variable and the lagged dependent variable, respectively, are defined as follows

$$\begin{aligned} \tilde{\mathbf{y}}_i &= (y_{i1} \quad y_{i2} \quad \cdots \quad y_{iT} \quad \Delta y_{i2} \quad \Delta y_{i3} \quad \cdots \quad \Delta y_{iT})' \\ \tilde{\mathbf{y}}_{i,-1} &= (y_{i0} \quad y_{i1} \quad \cdots \quad y_{i,T-1} \quad \Delta y_{i1} \quad \Delta y_{i2} \quad \cdots \quad \Delta y_{i,T-1})'. \end{aligned} \quad (31)$$

The matrix $\tilde{\mathbf{x}}_i$ for the exogenous variables is given by

$$\tilde{\mathbf{x}}_i = \begin{pmatrix} x_{i1}^1 & x_{i1}^2 & x_{i1}^3 & \cdots & x_{i1}^m \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ x_{iT}^1 & x_{iT}^2 & x_{iT}^3 & \cdots & x_{iT}^m \\ \Delta x_{i2}^1 & \Delta x_{i2}^2 & \Delta x_{i2}^3 & \cdots & \Delta x_{i2}^m \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \Delta x_{iT}^1 & \Delta x_{iT}^2 & \Delta x_{iT}^3 & \cdots & \Delta x_{iT}^m \end{pmatrix} \quad (32)$$

while the matrix $\tilde{\mathbf{w}}_i$ for the endogenous variables is defined as follows

$$\tilde{\mathbf{w}}_i = \begin{pmatrix} w_{i1}^1 & w_{i1}^2 & w_{i1}^3 & \cdots & w_{i1}^q \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ w_{iT}^1 & w_{iT}^2 & w_{iT}^3 & \cdots & w_{iT}^q \\ \Delta w_{i2}^1 & \Delta w_{i2}^2 & \Delta w_{i2}^3 & \cdots & \Delta w_{i2}^q \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \Delta w_{iT}^1 & \Delta w_{iT}^2 & \Delta w_{iT}^3 & \cdots & \Delta w_{iT}^q \end{pmatrix}. \quad (33)$$

Therefore $\hat{g}_N(\theta_0) = N^{-1} \sum_{i=1}^N G_i' \tilde{\mathbf{y}}_i - N^{-1} \sum_{i=1}^N G_i' \tilde{\mathbf{z}}_i \theta_0$. By Lemma 1, one may write the

likelihood for θ as

$$p \left(N^{-1} \sum_{i=1}^N G_i' \tilde{\mathbf{y}}_i \mid \theta, N^{-1} \sum_{i=1}^N G_i' \tilde{\mathbf{z}}_i \right) \propto \exp \left(-\frac{1}{2} N \hat{g}_N'(\theta) S_{(\theta)}^{-1} \hat{g}_N(\theta) \right). \quad (34)$$

Hence, the model likelihood can be expressed as

$$\int_{\Theta} p \left(N^{-1} \sum_{i=1}^N G_i' \tilde{\mathbf{y}}_i \mid \theta \right) p(\theta) d\theta \propto \int_{\Theta} \exp \left(-\frac{1}{2} N \hat{g}_N'(\theta) S_{(\theta)}^{-1} \hat{g}_N(\theta) \right) p(\theta) d\theta.$$

Assuming that the prior $p(\theta)$ is second order differentiable around $\hat{\theta}_0$ and using the Laplace approximation, we obtain that the model likelihood is proportional to

$$\int_{\Theta} p \left(N^{-1} \sum_{i=1}^N G_i' \tilde{\mathbf{y}}_i \mid \theta \right) p(\theta) d\theta \propto \exp \left(-\frac{1}{2} N \hat{g}_N'(\hat{\theta}_0) S_{(\hat{\theta}_0)}^{-1} \hat{g}_N(\hat{\theta}_0) + \log p(\hat{\theta}_0) + \frac{k}{2} \log 2\pi - \frac{1}{2} \log \det \frac{\partial^2}{\partial \theta^2} \left(\frac{1}{2} N \hat{g}_N'(\hat{\theta}_0) S_{(\hat{\theta}_0)}^{-1} \hat{g}_N(\hat{\theta}_0) \right) \right)$$

where $\hat{\theta}_0 \equiv \arg \min_{\theta} N \hat{g}_N'(\theta) S_{(\theta)}^{-1} \hat{g}_N(\theta)$ is the GMM estimate of θ_0 with weighting matrix $S_{(\theta)}^{-1}$. Noting the fact that $\partial^2 (\hat{g}_N' S^{-1} \hat{g}_N) / \partial \theta^2 |_{\theta=\hat{\theta}_0}$ is a $k \times k$ matrix of order $O_p(1)$ due to the ergodicity assumption, the model likelihood can be approximated by

$$\int_{\Theta} p \left(N^{-1} \sum_{i=1}^N G_i' \tilde{\mathbf{y}}_i \mid \theta \right) p(\theta) d\theta \propto \exp \left(-\frac{1}{2} N \hat{g}_N'(\hat{\theta}_0) S_{(\hat{\theta}_0)}^{-1} \hat{g}_N(\hat{\theta}_0) - \frac{k}{2} \log N \right) \quad (35)$$

where k is the dimension of vector θ . Alternatively, the above approximation has the order of $O_p(N^{-1/2})$ if the unit information prior for θ is used with $\partial^2 (\hat{g}_N' S^{-1} \hat{g}_N) / \partial \theta^2 |_{\theta=\hat{\theta}_0}$ as its variance-covariance matrix, that is, the prior distribution for θ , $p(\theta)$, is given by $N(0, \partial^2 (\hat{g}_N' S^{-1} \hat{g}_N) / \partial \theta^2 |_{\theta=\hat{\theta}_0})$.

For a given model M_j for which θ has k_j elements different from zero, with the estimate denoted by $\hat{\theta}_{0,j}$, the model likelihood (35) becomes

$$\int_{\Theta} p \left(N^{-1} \sum_{i=1}^N G_i' \tilde{\mathbf{y}}_i \mid \theta, M_j \right) p(\theta) d\theta \propto \exp \left(-\frac{1}{2} N \hat{g}_N'(\hat{\theta}_{0,j}) S_{(\hat{\theta}_{0,j})}^{-1} \hat{g}_N(\hat{\theta}_{0,j}) - \frac{k_j}{2} \log N \right). \quad (36)$$

Then the moment conditions (27) associated with model M_j can be written as

$$E \left[G_i' (\tilde{\mathbf{y}}_i - \tilde{\mathbf{z}}_i C_{M_j} \theta_0) \right] = 0 \quad \text{where } C_{M_j} \text{ is a diagonal choice matrix such that its diagonal will}$$

have 1's if the corresponding variable is included in the model and 0's otherwise.

Recognizing that the estimate $\hat{\theta}_0$ differs from model to model, the sample mean of the

moment conditions for model M_j can be written as $\hat{g}_N(\hat{\theta}_0) = N^{-1} \sum_{i=1}^N G_i' (\tilde{\mathbf{y}}_i - \tilde{\mathbf{z}}_i C_{M_j} \hat{\theta}_{0,j})$.

It is easy to see that G_i' , $\tilde{\mathbf{y}}_i$, and $\tilde{\mathbf{z}}_i$ are the same across all models. In other words, the

moment conditions and the observable data are the same across the universe of models.¹² allowing us to make valid comparisons of posterior probabilities, in accordance to the principle of Bayesian factor analysis. Therefore, by using (36), one can compute the posterior odds ratio of two models M_1 and M_2 by

$$\begin{aligned} \frac{p\left(M_1 \mid N^{-1} \sum_{i=1}^N G_i' \tilde{\mathbf{y}}_i\right)}{p\left(M_2 \mid N^{-1} \sum_{i=1}^N G_i' \tilde{\mathbf{y}}_i\right)} &= \frac{p(M_1)}{p(M_2)} \frac{p\left(N^{-1} \sum_{i=1}^N G_i' \tilde{\mathbf{y}}_i \mid M_1\right)}{p\left(N^{-1} \sum_{i=1}^N G_i' \tilde{\mathbf{y}}_i \mid M_2\right)} \\ &= \frac{p(M_1)}{p(M_2)} \exp\left(-\frac{1}{2}\left(N\hat{g}_N'(\hat{\theta}_{0,1})S_{(\hat{\theta}_{0,1})}^{-1}\hat{g}_N(\hat{\theta}_{0,1}) - N\hat{g}_N'(\hat{\theta}_{0,2})S_{(\hat{\theta}_{0,2})}^{-1}\hat{g}_N(\hat{\theta}_{0,2})\right) - \left(\frac{k_1 - k_2}{2} \log N\right)\right), \end{aligned} \quad (37)$$

which has the same form of BIC as fully specified models. We use iterative GMM estimation with moment conditions $E\left[G_i'(\tilde{\mathbf{y}}_i - \tilde{\mathbf{z}}_i C_{M_j} \theta_{0,j})\right] = 0$ to approximate the Bayesian factors above. A consistent estimate of the weighting matrix is used to replace $S_{(\hat{\theta}_0)}^{-1}$ in (37).

IV. MONTE CARLO SIMULATION AND RESULTS

In this section we describe the Monte Carlo simulations intended to assess the performance of LIBMA. We compute posterior model probabilities, inclusion probabilities for each variable in the universe considered, and parameter statistics. These statistics provide a description of how well our procedure helps the inference process both in a Bayesian model selection and a Bayesian model averaging framework.

A. The Data Generating Process

We consider the case where the universe of potential explanatory variables contains 6 exogenous variables, 2 endogenous variables and the lagged dependent variable. Throughout our simulations we maintain the number of periods constant, that is, $T = 4$ and we vary the number of individuals, N .

¹² This approach is in line with the model selection procedure proposed by Andrews and Lu (2001).

For every individual i and period t , the first four exogenous variables are generated as follows

$$\begin{pmatrix} x_{it}^1 & x_{it}^2 & x_{it}^3 & x_{it}^4 \end{pmatrix} = \begin{pmatrix} 0.3 & 0.4 & 0.8 & 0.5 \end{pmatrix} + \mathbf{r}_t \quad (38)$$

with $\mathbf{r}_t \sim N(0, I_4)$ for $t = 0, 1, \dots, T$; $i = 1, \dots, N$,

where I_4 is the four dimensional identity matrix. We allow for some correlation between the first two and the last two exogenous variables. That is, $\begin{pmatrix} \mathbf{x}_i^5 & \mathbf{x}_i^6 \end{pmatrix}$ are correlated with $\begin{pmatrix} \mathbf{x}_i^1 & \mathbf{x}_i^2 \end{pmatrix}$ such that for every individual i and period t , the data generating process is given by

$$\begin{pmatrix} x_{it}^5 & x_{it}^6 \end{pmatrix} = \left(\begin{pmatrix} x_{it}^1 & x_{it}^2 \end{pmatrix} - \begin{pmatrix} 0.3 & 0.4 \end{pmatrix} \right) \cdot 0.1 \cdot \begin{pmatrix} 1 & 2 \end{pmatrix}' \begin{pmatrix} 1 & 1 \end{pmatrix} + \begin{pmatrix} 1.5 & 1.8 \end{pmatrix} + \mathbf{r}_t \quad (39)$$

with $\mathbf{r}_t \sim N(0, I_2)$ for $t = 0, 1, \dots, T$; $i = 1, \dots, N$,

where I_2 is the two dimensional identity matrix.

Similarly, for the endogenous variables, $\begin{pmatrix} \mathbf{w}_i^1 & \mathbf{w}_i^2 \end{pmatrix}$, we have the following data generating process

$$\begin{aligned} \begin{pmatrix} w_{it}^1 & w_{it}^2 \end{pmatrix} &= 0.9 \begin{pmatrix} w_{i,t-1}^1 & w_{i,t-1}^2 \end{pmatrix} + 10v_{it} \begin{pmatrix} 1 & 1 \end{pmatrix} + \mathbf{r}_t \quad \text{for } t = 1, 2, \dots, T \\ \begin{pmatrix} w_{i0}^1 & w_{i0}^2 \end{pmatrix} &= 10v_{i0} \begin{pmatrix} 1 & 1 \end{pmatrix} + \mathbf{r}_0 \\ \text{with } v_{it} &\sim N(0, \sigma_v^2) \quad \text{and } \mathbf{r}_t \sim N(0, I_2) \quad \text{for } t = 0, 1, \dots, T. \end{aligned} \quad (40)$$

As the data generating process for the endogenous variables indicates, the overall error term v_{it} is assumed to be distributed normally here. We relax the normality assumption later.

For $t = 0$, the dependent variable is generated by

$$\begin{aligned} y_{i0} &= \frac{1}{(1 - \alpha)} (\mathbf{x}_{i0} \theta_x + \mathbf{w}_{i0} \theta_w + \eta_i + v_{i0}) \\ \text{with } v_{i0} &\sim N(0, \sigma_v^2) \quad \text{and } \eta_i \sim N(0, \sigma_\eta^2) \end{aligned} \quad (41)$$

where $\theta_x = \begin{pmatrix} 0.07 & 0 & 0 & -0.09 & 0 & 0.1 \end{pmatrix}'$, $\theta_w = \begin{pmatrix} 0 & -0.1 \end{pmatrix}'$, $\mathbf{w}_{i0} = \begin{pmatrix} w_{i0}^1 & w_{i0}^2 \end{pmatrix}$, and $x_{i0} = \begin{pmatrix} x_{i0}^1 & x_{i0}^2 & x_{i0}^3 & x_{i0}^4 & x_{i0}^5 & x_{i0}^6 \end{pmatrix}$.

For $t = 1, 2, \dots, T$ the data generating process is given by

$$\begin{aligned} y_{it} &= \alpha y_{i,t-1} + \theta_x \mathbf{x}_{it} + \theta_w \mathbf{w}_{it} + \eta_i + v_{it} \\ \text{with } v_{it} &\sim N(0, \sigma_v^2) \quad \text{and } \eta_i \sim N(0, \sigma_\eta^2). \end{aligned} \quad (42)$$

We now test the robustness of our procedure with respect to underlying distributions of the error term by relaxing the normality assumption and using discrete distributions instead.

Concretely, to set the distribution of the random variable v_{it} , we first generate its support, S_v , by taking N_v points from a uniform sampling over the interval $[-1,1]$. Then we draw N_v i.i.d. random variables $\omega_k \sim \text{Exponential}(1)$. The probability mass assigned to each point $s_k \in S_v$ is obtained by setting $p_k = \frac{\omega_k}{\sum_i \omega_i}$. Finally, we translate each point in S_v so that v_{it} has zero mean. It is well known that the probability distribution obtained in this fashion is equivalent to a uniform sampling from a simplex in N_v dimensional space. The construction of the simulated model follows exactly the case of the Normal distribution, with the only difference being the use of the discrete distribution described above in every place where the Normal distribution is used for v_{it} .

B. Simulation Results

This section reports Monte Carlo simulations of our LIBMA methodology in order to assess its performance. We generate 100 instances of the data generating process with the exogenous variables \mathbf{x}_{it} , endogenous variables \mathbf{w}_{it} , and parameter values $(\alpha \ \theta_x \ \theta_w)'$ as discussed in the previous section, and we present results in the form of medians, means, variances and quartiles. We consider several sample sizes, $N = 200, 500, 1000$ and 2000 , and several values for the coefficient of the lagged dependent variable, $\alpha = 0.95, 0.50$, and 0.30 . In the first set of simulations we assume that both the random error term v_{it} and the individual effect η_i are drawn from a Normal distribution, $v_{it} \sim N(0, \sigma_v^2)$ and $\eta_i \sim N(0, \sigma_\eta^2)$, respectively. We consider the cases where $\sigma_v = 0.05, 0.10$, and 0.50 while $\sigma_\eta = 0.10$. Since our methodology should not depend on the normality of the random error term, we check for robustness by creating a second set of simulations where the assumption of normality for v_{it} is dropped, as discussed earlier.

Model selection

In the Bayesian framework, the posterior model probability is a key indicator of performance. Table 1 presents means, variances, and three quartiles (Q1, median, and Q3) for the posterior probability of the true model across the 100 instances. As expected, the mean posterior probabilities of the true model increase with the sample size. For sample size values of $N = 200, 500, 1000$ and 2000 , average values of the posterior model probability

are about 0.31, 0.46, 0.57, and 0.65, respectively. Median posterior model probabilities are slightly higher than the means, with average values of 0.32, 0.50, 0.62, and 0.69. In addition, as the sample increases, the distribution becomes skewed toward 1. Quartiles and distribution plots show that as the sample increases the distributions of the posterior model probabilities are becoming less and less normal, with long left tails.¹³

As shown in (3) the posterior model probability depends on the prior model probability. Under the assumption that all models have equal prior probability, the more variables are under consideration the smaller the prior probability for each model. Obviously that has an effect on the absolute value of the posterior model probability. Therefore, we choose to also compute a relative measure that helps one understand how well the methodology performs. Table 2 presents the ratio of the posterior model probability of the true model to the highest posterior probability of all the other models (excluding the true model). On average this ratio is above unity for all the cases considered, suggesting that the correct model is on average favored over all the other models. As expected, the average ratios increase with the sample size, starting from about 2.26 for $N = 200$ and reaching 7.09 for $N = 2000$.

In Table 3 we examine how often our methodology recovers the true model by reporting how many times, out of 100 instances, the true model has the highest posterior probability. The results indicate that this is done quite well. For the smallest sample size, $N = 200$, the recovery rate varies from 65 percent to 83 percent. For $N = 500$ we see an improvement in the selection of the true model with the success rate ranging from 82 percent to 93 percent. For sample sizes bigger than 1000, the recovery rate stays over 90 percent, reaching 97 percent in a couple of cases.

Model averaging

While model selection properties are desired, researchers are often more interested in making inferences. Table 4 presents the posterior inclusion probabilities for all the variables considered along with the true model (column 2 of the table).¹⁴ Given the assumptions made relative to the model priors, the prior probability of inclusion for each variable is the same and equal to 0.5. From Table 4 we see that the median value of the inclusion probability for all the relevant explanatory variables is greater than 0.942 in all cases considered. As the

¹³ Figures 1, 2, 4, and 5 in Appendix A show density plots for the posteriors in Tables 1, 2, 6, and 7.

¹⁴ A value of 1(0) in column 2 indicates that the true model contains (excludes) that variable.

sample size increases the posterior inclusion probabilities approach 1 for all the relevant variables. In fact for sample sizes greater than 500, the median value of the probability of inclusion for all relevant variables is practically 1. For the variables not contained in the true model the median posterior probability of inclusion decreases with the sample size with the upper bound being less than 0.076 for the case when $N = 2000$.

We turn now to the parameter estimates, and examine how the estimated values compare with the true parameter values. Table 5 presents the median values of the estimated parameters, averaged over 100 replications, compared to the parameters of the true model.¹⁵ As in the case of inclusion probabilities, our methodology is performing very well in estimating the parameters, with the performance improving as the sample gets larger. In Figure 3 of Appendix A, we present the box plots for the parameter estimates of Table 5, for the case of $\alpha = 0.95$ and $\sigma_v = 0.1$. It becomes clear that as the sample increases the variance of the distribution decreases and the median converges to the true value. Aside from the fact that the estimates are very close to the true parameter values, the variance over the 100 replications is also very small across the board with values less than 10^{-5} in many cases.

Robustness checks using non-Gaussian errors

As discussed in Section A we perform robustness analysis by relaxing the normality assumption for the error term v_{it} . Overall, as shown in Tables 6-10, the results are very similar to those presented in Tables 1-5. Tables 6 and 7 (which are analogous to Tables 1 and 2), present posterior model probabilities for the true model, and the ratio of the posterior model probability of the true model to the highest posterior probability of all other models, respectively. In Table 6, average mean and median values of the posterior model probability are about the same as in Table 1. Median posterior model probabilities are again slightly higher than the means, a result that is consistent with a skewed distribution. In addition, as the sample increases, the posterior model probabilities increase and the distribution becomes increasingly more skewed toward 1. For Table 7, conclusions are similar to Table 2. On average, the ratio of the posterior model probability of the true model to the highest posterior probability of all other models is above unity for all cases considered

¹⁵ Parameter values are discussed in section 4. Essentially these are constant for $x_1, x_2, x_3, x_4, x_5, x_6, w_1$ and w_2 , and vary for y_{t-1} based on the values of $\alpha = 0.95, 0.50$ and 0.30 .

suggesting that the correct model is, on average, favored over all the other models. As expected, the average ratios increase with the sample size, starting from about 1.66 for $N = 200$, and reaching 7.29 for $N = 2000$.

Model recovery under non-Gaussian errors is still very good. As shown in Table 8, results are very similar to those of Table 3. For the smallest sample size, $N = 200$, the recovery rate varies from 65 percent to 82 percent. For $N = 500$ we see an improvement in the selection of the true model with the success rate ranging from 81 percent to 91 percent. For the biggest sample size considered, $N = 2000$, the recovery rate stays over 88 percent, reaching again 97 percent for $\alpha = 0.30$.

Tables 9 and 10 present the posterior inclusion probabilities and parameter estimates using LIBMA and compares them the true model. From Table 9, we see that, the median inclusion probability for all the relevant explanatory variables is greater than 0.939 in all cases considered. As the sample size increases, the posterior inclusion probabilities approach 1 for all the relevant variables. As in the Gaussian case, for sample sizes greater than 500, the probability of inclusion for all relevant variables is practically 1. For the variables not contained in the true model the median posterior probability of inclusion decreases with the sample size with the upper bound being less than 0.078 for the case when $N = 2000$. In Table 10, estimated parameter medians and variances are almost identical to those reported in Table 5. As in the Gaussian case, our methodology performs well by generating estimates close to the true parameter values for all types of variables, with performance improving with larger samples. In Figure 6 of Appendix A the box plots for the parameter estimates of Table 10 (for the case of $\alpha = 0.95$ and $\sigma_v = 0.1$) indicate that the variance of the distribution decreases and the median converges to the true value asymptotically.

V. CONCLUSION

This paper proposes a limited information methodology in the context of Bayesian Model Averaging, which we label LIBMA, for panel data models where the lagged dependent variable appears as a regressor and endogenous variables are present. The LIBMA methodology incorporates a GMM estimator for dynamic panel data models in a Bayesian Model Averaging framework to explicitly account for model uncertainty. Our methodology adds value to the existing literature in three important ways. First, while standard BMA is a full information technique where a complete stochastic specification is assumed, LIBMA is a limited information approach which relies on GMM estimation, a limited information

technique based on moment restrictions rather than a complete stochastic specification. Second, LIBMA explicitly controls for endogeneity. The likelihood and exact expressions of the marginal likelihood used in the fully Bayesian analyses are replaced by the limited information construct modeled on the GMM estimation, and a limited information criterion as an approximation to the actual marginal likelihoods, respectively. Third, we use this methodology in a panel setting thus expanding its usability to a wide range of applications.

Based on simulation results, we conclude that asymptotically LIBMA performs very well and it can be used to address the issue of model uncertainty in the dynamic panel data models with endogenous regressors.

Future research should explore the possibility of using the LIBMA methodology for applications where the sample size is constrained by data availability, such as those investigating robust patterns of cross-country growth behavior.

REFERENCES

- Andrews, D. W. K., and B. Lu, 2001, "Consistent Model and Moment Selection Procedures for GMM Estimation with Application to Dynamic Panel Data Models," *Journal of Econometrics*, Vol. 101, No. 1, pp. 123–64.
- Aragones, E., I. Gilboa, A. Postlewaite, and D. Schmeidler, 2005, "Fact-Free Learning," *American Economic Review* Vol. 95, No. 5, pp. 1355–68.
- Back, K. and D. P. Brown, 1993, "Implied Probabilities in GMM Estimators", *Econometrica*, Vol. 61, No. 4, pp. 971–975.
- Blundell, R., and S. Bond, 1998, "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, Vol. 87, No. 1, pp. 114–43.
- Brock, W., and S. Durlauf, 2001, "Growth Empirics and Reality," *World Bank Economic Review*, 15, pp. 229–272.
- Chernozhukov, V., and H. Hong, 2003, "An MCMC Approach to Classical Estimation," *Journal of Econometrics*, Vol. 115, No. 2, pp. 293–346.
- Chipman, H., E.I. George, and R.E. McCulloch, 2001, "The Practical Implementation of Bayesian Model Selection," (with discussion) in P. Lahiri (ed), *Model Selection* IMS Lecture Notes, Vol. 38, pp. 70–134.
- Durlauf, S., and D. Quah, 1999, "The New Empirics of Economic Growth," in J. B. Taylor and M. Woodford (eds), *Handbook of Macroeconomics* Vol. IA (North Holland).
- Fernàndez C., E. Ley, and M. Steel, 2001a, "Model Uncertainty in Cross-Country Growth Regressions," *Journal of Applied Econometrics*, Vol. 16, pp. 563–76.
- Fernàndez C., E. Ley and M.F.J. Steel, 2001b, "Benchmark Priors for Bayesian Model Averaging," *Journal of Econometrics*, Vol. 100, pp. 381–427.
- Hall, A.R., 2005, *Generalized Method of Moments* (New York: Oxford University Press).
- Hoeting, J.A., D. Madigan, A.E. Raftery, and C.T. Volinsky, 1999, "Bayesian Model Averaging: A Tutorial," *Statistical Science*, Vol. 14, No. 4, pp. 382–417.
- Hong, H., and B. Preston, 2008, "Bayesian Averaging, Prediction and Nonnested Model Selection," *NBER Working Papers 14284*, National Bureau of Economic Research, Inc.
- Jacobson T., and S. Karlsson, 2004, "Finding Good Predictors for Inflation: A Bayesian Model Averaging Approach", *Journal of Forecasting*, Vol. 23, pp. 476–496.
- Kass, R., and A. Raftery, 1995, "Bayes Factors," *Journal of the American Statistical Association*, Vol. 90, No. 430, pp. 773–95.

- Kass, R., and L. Wasserman, 1995, "A Reference Bayesian Test for Nested Hypotheses and Its Relationship to the Schwarz Criterion," *Journal of the American Statistical Association*, Vol. 90, No. 431, pp. 928–34.
- Kim, J. Y., 2002, "Limited Information Likelihood and Bayesian Analysis," *Journal of Econometrics*, Vol. 107, No. 1-2, pp. 175–93.
- Koop, G., and L. Tole, 2004, "Measuring the Health Effects of Air Pollution: to What Extent Can We Really Say that People are Dying from Bad Air?" *Journal of Environmental Economics and Management* 47, January, pp. 30–54.
- Leamer, E., 1978, *Specification Searches: Ad Hoc Inference with Non-experimental Data* (New York: Wiley).
- Leamer, E., 1983, "Let's Take the Con Out of Econometrics," *American Economic Review*, Vol. 73, pp. 31–43.
- Ley E., and M. Steel, 2008, "On the Effect of Prior Assumptions in Bayesian Model Averaging with Applications to Growth Regression," *Journal of Applied Econometrics*, forthcoming.
- Madigan, D.M. and Raftery, A.E., 1994, "Model Selection and Accounting for Model Uncertainty in Graphical Models using Occam's Window," *Journal of the American Statistical Association*, Vol. 89, pp. 1335–1346.
- Mitchell, T.J., and J.J. Beauchamp, 1988, "Bayesian Variable Selection in Linear Regression," *Journal of the American Statistical Association* Vol. 83, pp. 1023–1032.
- Moral-Benito, E., 2007, "Determinants of Economic Growth: A Bayesian Panel Data Approach," CEMFI Working Papers Paper 0719.
- Morales, K.H., J.G. Ibrahim, C. Chen, and L.M. Ryan, 2006, "Bayesian Model Averaging With Applications to Benchmark Dose Estimation for Arsenic in Drinking Water," *Journal of the American Statistical Association* 101, Vol. 473, pp. 9–17.
- Moulton, B.R., 1991, "A Bayesian Approach to Regression Selection and Estimation with Application to a Price Index for Radio Services," *Journal of Econometrics*, Vol. 49, pp. 169–93.
- Raftery, A. E., 1995, "Bayesian Model Selection in Social Research," *Sociological Methodology*, Vol. 25, pp. 111–163.
- Raftery, A. E., 1996, "Approximate Bayes Factors and Accounting for Model Uncertainty in Generalized Linear Models," *Biometrika*, Vol. 83, pp. 251–66.
- Raftery, A.E., D. Madigan, and J. A. Hoeting, 1997, "Bayesian Model Averaging for Linear Regression Models," *Journal of the American Statistical Association*, 92, pp. 179–191.
- Ragusa, G., 2008, "Bayesian Likelihoods for Moment Condition Models," mimeo.

- Sala-i-Martin, X., Doppelhofer, G., R., and I. Miller, 2004, "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach," *American Economic Review*, Vol. 94, No. 4, pp. 813–35.
- Schennach, S.C., 2005, "Bayesian Exponentially Tilted Empirical Likelihood," *Biometrika*, Vol. 92, pp. 31–46.
- Tsangarides, C., 2004, "A Bayesian Approach to Model Uncertainty," IMF Working Paper No. 04/68 (Washington: International Monetary Fund).
- Wintle, B.A., M.A. McCarthy, C.T. Volinsky, and R.P. Kavanagh, 2003, "The Use of Bayesian Model Averaging to Better Represent Uncertainty in Ecological Models," *Conservation Biology* 17, December, pp. 1579–1590.
- Yeung, K.Y., R.E. Bumgarner, and A. E. Raftery, 2005, "Bayesian Model Averaging: Development of an Improved Multi-Class, Gene Selection and Classification Tool for Micro array Data," *Bioinformatics* 21, Vol. 10, pp. 2394–2402.

Table 1. Posterior Probability of the True Model
Summary statistics using LIBMA estimation for various n, a, and s_v

	a	0.95			0.50			0.30		
		0.05	0.10	0.50	0.05	0.10	0.50	0.05	0.10	0.50
<i>Sample</i>										
<i>N=200</i>										
Mean		0.318	0.328	0.286	0.341	0.331	0.258	0.326	0.333	0.255
Variance		0.017	0.013	0.013	0.015	0.018	0.018	0.018	0.014	0.014
Q1		0.231	0.255	0.202	0.276	0.248	0.156	0.255	0.255	0.167
Median		0.339	0.327	0.294	0.361	0.355	0.268	0.337	0.357	0.258
Q3		0.425	0.430	0.386	0.426	0.440	0.363	0.439	0.425	0.324
<i>N=500</i>										
Mean		0.487	0.465	0.440	0.470	0.459	0.450	0.468	0.458	0.474
Variance		0.017	0.021	0.021	0.023	0.024	0.019	0.022	0.023	0.020
Q1		0.426	0.378	0.336	0.449	0.378	0.363	0.391	0.339	0.417
Median		0.519	0.501	0.467	0.523	0.501	0.475	0.493	0.502	0.528
Q3		0.576	0.578	0.562	0.567	0.577	0.559	0.592	0.570	0.579
<i>N=1000</i>										
Mean		0.567	0.552	0.561	0.561	0.570	0.562	0.588	0.577	0.586
Variance		0.026	0.025	0.021	0.023	0.024	0.022	0.018	0.026	0.021
Q1		0.509	0.473	0.498	0.490	0.513	0.489	0.545	0.543	0.525
Median		0.622	0.602	0.611	0.605	0.632	0.604	0.632	0.623	0.628
Q3		0.679	0.668	0.666	0.671	0.685	0.676	0.678	0.682	0.698
<i>N=2000</i>										
Mean		0.646	0.650	0.667	0.638	0.636	0.645	0.633	0.636	0.670
Variance		0.020	0.019	0.014	0.027	0.025	0.027	0.021	0.018	0.015
Q1		0.605	0.570	0.614	0.573	0.549	0.611	0.567	0.574	0.641
Median		0.690	0.698	0.705	0.704	0.686	0.701	0.680	0.655	0.700
Q3		0.748	0.747	0.750	0.740	0.748	0.752	0.738	0.736	0.757

Notes:

1. For the idiosyncratic error term, $\eta_i \sim N(0, \sigma_\eta^2)$ where $\sigma_\eta = 0.10$.
2. The error term is normally distributed $v_{it} \sim N(0, \sigma_v^2)$.

Table 2. Posterior Probability Ratio of True Model/Best Among the Other Models
Summary statistics using LIBMA estimation for various n, α , and σ_v

Sample	α	0.95			0.50			0.30		
		σ_v			σ_v			σ_v		
		0.05	0.10	0.50	0.05	0.10	0.50	0.05	0.10	0.50
<i>N=200</i>										
Mean		2.044	2.102	2.061	2.255	2.201	1.809	2.117	2.172	1.681
Variance		1.932	1.627	1.655	1.773	2.089	1.709	1.928	1.740	1.332
Q1		0.902	1.118	0.921	1.339	0.980	0.689	1.007	1.027	0.710
Median		1.715	1.810	1.848	2.031	1.991	1.611	1.927	2.010	1.546
Q3		3.197	3.015	2.968	3.046	3.276	2.609	3.317	3.021	2.483
<i>N=500</i>										
Mean		3.729	3.486	3.151	3.456	3.372	3.094	3.583	3.399	3.594
Variance		4.285	4.292	4.405	4.083	4.823	3.583	5.232	4.795	4.126
Q1		2.145	1.515	1.568	2.066	1.535	1.646	1.607	1.623	1.950
Median		3.538	3.479	2.692	3.377	3.282	2.921	3.078	3.186	3.537
Q3		5.063	4.998	4.409	4.815	4.901	4.476	5.496	4.644	5.185
<i>N=1000</i>										
Mean		5.174	4.797	4.732	4.809	5.155	4.911	5.420	5.208	5.621
Variance		10.470	9.862	7.615	9.301	9.078	9.004	9.109	8.798	11.017
Q1		2.811	2.098	2.380	2.146	2.318	2.163	3.189	2.802	2.819
Median		4.754	4.363	4.682	4.530	5.080	4.924	5.211	4.895	5.743
Q3		7.167	7.065	6.629	6.744	7.530	7.233	7.588	7.644	8.412
<i>N=2000</i>										
Mean		6.479	6.808	7.088	6.777	6.592	6.971	6.240	6.200	6.881
Variance		14.244	19.297	15.844	18.505	18.891	18.147	15.645	17.566	15.194
Q1		3.144	3.526	3.811	3.035	2.826	3.642	3.159	2.873	4.086
Median		6.616	6.500	6.892	6.597	5.822	6.714	5.359	5.223	6.078
Q3		9.657	8.884	10.054	9.258	9.360	10.253	9.033	8.732	9.433

Notes:

1. For the idiosyncratic error term, $\eta_i \sim N(0, \sigma_\eta^2)$ where $\sigma_\eta = 0.10$.
2. The error term is normally distributed $v_{it} \sim N(0, \sigma_v^2)$.

Table 3. Probability of Retrieving the True Model
Summary statistics using LIBMA estimation for various n, α , and σ_v

<i>Sample</i>	α	0.95			0.50			0.30		
		σ_v			σ_v			σ_v		
		0.05	0.10	0.50	0.05	0.10	0.50	0.05	0.10	0.50
<i>N=200</i>										
% Correct		74	78	74	83	73	65	76	76	65
<i>N=500</i>										
% Correct		91	87	83	83	84	83	85	82	87
<i>N=1000</i>										
% Correct		91	92	90	90	91	91	91	92	90
<i>N=2000</i>										
% Correct		94	93	97	92	94	92	92	94	97

Notes:

1. For the idiosyncratic error term, $\eta_i \sim N(0, \sigma_\eta^2)$ where $\sigma_\eta = 0.10$.
2. The error term is normally distributed $v_{it} \sim N(0, \sigma_v^2)$.

Table 4. Model Recovery: Medians and Variances of Posterior Inclusion Probability for Each Variable
True model vs BMA posterior inclusion probability for various n , α , and σ_v

Sample	True Model	α						σ_v							
		0.95			0.50			0.10			0.50				
		Median	Variance	Median	Variance	Median	Variance	Median	Variance	Median	Variance	Median	Variance		
$N=200$															
y_{t-1}	1	0.9999	0.0007	1.0000	0.00842	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	0.9991	0.00132
x_1	1	1.0000	0.0005	1.0000	0.00278	0.96488	0.03435	1.0000	0.0000	1.0000	0.0000	0.96058	0.04162	1.0000	0.05247
x_2	0	0.1808	0.02581	0.18371	0.00990	0.17545	0.00963	0.18805	0.01379	0.18221	0.01599	0.18153	0.01524	0.17765	0.01558
x_3	0	0.17971	0.02050	0.19415	0.01944	0.18508	0.00967	0.17270	0.00808	0.17759	0.01156	0.18634	0.01762	0.16574	0.01977
x_4	1	1.0000	0.0000	1.0000	0.00166	0.99705	0.01167	1.0000	0.0000	1.0000	0.0000	0.99676	0.00448	1.0000	0.00878
x_5	0	0.17414	0.02745	0.18618	0.01550	0.19205	0.02060	0.17705	0.01156	0.19663	0.02901	0.19163	0.03022	0.19357	0.01167
x_6	1	1.0000	0.0002	1.0000	0.00007	0.99954	0.01189	1.0000	0.0000	1.0000	0.0000	0.99956	0.00642	1.0000	0.00107
w_1	0	0.18608	0.02415	0.18098	0.01545	0.18245	0.01185	0.20964	0.02637	0.17619	0.02792	0.20706	0.02546	0.19416	0.02392
w_2	1	1.0000	0.0001	1.0000	0.00048	0.99700	0.00342	1.0000	0.0000	1.0000	0.0000	0.99912	0.00774	1.0000	0.00402
$N=500$															
y_{t-1}	1	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000
x_1	1	1.0000	0.0000	1.0000	0.0000	0.99998	0.00234	1.0000	0.0000	1.0000	0.0000	0.99999	0.00004	1.0000	0.00102
x_2	0	0.11471	0.00562	0.12409	0.01272	0.12373	0.01309	0.11931	0.02216	0.12556	0.02333	0.13176	0.00982	0.12793	0.01339
x_3	0	0.12296	0.02203	0.12361	0.02297	0.11970	0.00932	0.11993	0.00931	0.12440	0.01291	0.11679	0.01591	0.13081	0.01963
x_4	1	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000
x_5	0	0.12323	0.00889	0.12552	0.01465	0.14116	0.01515	0.12201	0.02015	0.11921	0.02389	0.13090	0.01843	0.11717	0.01003
x_6	1	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000
w_1	0	0.12367	0.00938	0.12453	0.01410	0.12666	0.02661	0.12565	0.02103	0.12256	0.00872	0.13200	0.01786	0.12191	0.01421
w_2	1	1.0000	0.0000	1.0000	0.0000	1.0000	0.0001	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000
$N=1000$															
y_{t-1}	1	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000
x_1	1	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000
x_2	0	0.08121	0.01276	0.09082	0.01233	0.09906	0.01312	0.08616	0.01047	0.09000	0.01126	0.08321	0.00340	0.08698	0.01058
x_3	0	0.09067	0.01760	0.08997	0.01588	0.08408	0.00318	0.08956	0.00524	0.09266	0.01912	0.10138	0.01361	0.08588	0.01081
x_4	1	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000
x_5	0	0.08527	0.01763	0.09791	0.01669	0.10089	0.01323	0.08815	0.02364	0.08888	0.00658	0.09657	0.02133	0.08928	0.00590
x_6	1	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000
w_1	0	0.09031	0.01241	0.08731	0.01285	0.08880	0.01844	0.09128	0.01418	0.08429	0.01703	0.08349	0.01242	0.09136	0.01380
w_2	1	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000
$N=2000$															
y_{t-1}	1	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000
x_1	1	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000
x_2	0	0.06763	0.00546	0.05985	0.00760	0.06161	0.00284	0.06243	0.00929	0.06556	0.00321	0.06962	0.00456	0.07161	0.00171
x_3	0	0.07427	0.01254	0.06841	0.00778	0.06599	0.00295	0.07241	0.02087	0.07633	0.01132	0.06695	0.00862	0.06693	0.00407
x_4	1	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000
x_5	0	0.06433	0.00377	0.06468	0.01026	0.07099	0.00828	0.06543	0.00697	0.06466	0.02170	0.07229	0.02301	0.06856	0.00161
x_6	1	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000
w_1	0	0.06548	0.01370	0.06354	0.01062	0.06229	0.01246	0.06653	0.01232	0.06183	0.00823	0.06197	0.01229	0.07061	0.00968
w_2	1	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000

Notes:

1. For the idiosyncratic error term, $\eta_i \sim N(0, \sigma_\eta^2)$ where $\sigma_\eta = 0.10$.
2. The error term is normally distributed $v_{it} \sim N(0, \sigma_v^2)$.

Table 5. Model Recovery: Medians and Variances of Estimated Parameter Values
True model vs BMA coefficients' estimated values for various n, α , and σ_v .

α	0.95						0.50						0.30					
	0.05			0.10			0.05			0.10			0.05			0.10		
	True Value	Median	Variance	True Value	Median	Variance	True Value	Median	Variance	True Value	Median	Variance	True Value	Median	Variance	True Value	Median	Variance
$N=2000$																		
y_{t-1}	0.95	0.95080	0.00008	0.94952	0.00760	0.00113	0.50	0.50827	0.00053	0.52082	0.00108	0.52322	0.00161	0.30	0.31311	0.00039	0.31615	0.00130
x_1	0.07	0.06998	0.00001	0.06966	0.00006	0.00076	0.07	0.07048	0.00001	0.07041	0.00002	0.06702	0.00080	0.07	0.07052	0.00001	0.07211	0.00002
x_2	0.00	0.00005	0.00000	-0.00008	0.00001	0.00031	0.00	-0.00001	0.00000	0.00005	0.00000	-0.00030	0.00009	0.00	0.00003	0.00000	0.00003	0.00000
x_3	0.00	-0.00002	0.00000	-0.00001	0.00001	0.00029	0.00	-0.00006	0.00000	0.00006	0.00000	0.00024	0.00009	0.00	0.00003	0.00000	0.00000	0.00001
x_4	-0.09	-0.08964	0.00001	-0.09001	0.00016	0.00054	-0.09	-0.08968	0.00001	-0.09060	0.00002	-0.08967	0.00052	-0.09	-0.09024	0.00001	-0.09024	0.00002
x_5	0.00	0.00007	0.00000	0.00017	0.00000	0.00111	0.00	-0.00002	0.00000	-0.00012	0.00001	-0.00003	0.00018	0.00	0.00004	0.00000	-0.00003	0.00000
x_6	0.10	0.09986	0.00001	0.09850	0.00007	0.10068	0.10	0.09995	0.00001	0.10071	0.00002	0.10218	0.00055	0.10	0.10072	0.00001	0.10174	0.00002
w_1	0.00	0.00027	0.00001	0.00049	0.00002	-0.00023	0.00	0.00021	0.00001	0.00024	0.00001	0.00037	0.00022	0.00	0.00041	0.00002	0.00019	0.00001
w_2	-0.10	-0.09808	0.00005	-0.09688	0.00043	-0.09837	-0.10	-0.09601	0.00003	-0.09639	0.00004	-0.09725	0.00024	-0.10	-0.09704	0.00003	-0.09407	0.00004
$N=5000$																		
y_{t-1}	0.95	0.95075	0.00001	0.95015	0.00001	0.00039	0.50	0.50641	0.00020	0.50444	0.00047	0.50665	0.00066	0.30	0.30422	0.00018	0.31268	0.00041
x_1	0.07	0.06961	0.00000	0.06961	0.00001	0.00023	0.07	0.07019	0.00000	0.07040	0.00001	0.07125	0.00015	0.07	0.07011	0.00000	0.07051	0.00001
x_2	0.00	0.00000	0.00000	0.00005	0.00000	0.00002	0.00	-0.00002	0.00000	0.00002	0.00000	0.00031	0.00002	0.00	0.00002	0.00000	-0.00001	0.00000
x_3	0.00	0.00002	0.00000	0.00002	0.00000	0.00002	0.00	-0.00005	0.00000	-0.00010	0.00000	0.00016	0.00003	0.00	-0.00003	0.00000	0.00001	0.00000
x_4	-0.09	-0.09028	0.00000	-0.09081	0.00001	-0.08947	-0.09	-0.08979	0.00000	-0.08996	0.00001	-0.09025	0.00015	-0.09	-0.09015	0.00000	-0.09030	0.00001
x_5	0.00	0.00003	0.00000	0.00001	0.00000	-0.00010	0.00	-0.00004	0.00000	0.00000	0.00000	0.00033	0.00003	0.00	0.00000	0.00000	-0.00015	0.00000
x_6	0.10	0.10006	0.00000	0.09882	0.00001	0.10229	0.10	0.09978	0.00000	0.10015	0.00001	0.10217	0.00012	0.10	0.10050	0.00000	0.10020	0.00001
w_1	0.00	0.00008	0.00000	0.00013	0.00000	0.00007	0.00	0.00003	0.00000	0.00010	0.00000	0.00036	0.00003	0.00	0.00014	0.00000	0.00012	0.00000
w_2	-0.10	-0.09916	0.00002	-0.09802	0.00002	-0.09900	-0.10	-0.09732	0.00001	-0.09895	0.00001	-0.09886	0.00004	-0.10	-0.09840	0.00002	-0.09575	0.00002
$N=1000$																		
y_{t-1}	0.95	0.94991	0.00000	0.95014	0.00001	0.00024	0.50	0.50367	0.00007	0.50068	0.00020	0.50625	0.00023	0.30	0.30355	0.00010	0.30899	0.00023
x_1	0.07	0.06989	0.00000	0.06980	0.00000	0.00008	0.07	0.07011	0.00000	0.06978	0.00000	0.07113	0.00008	0.07	0.07023	0.00000	0.07034	0.00000
x_2	0.00	0.00000	0.00000	-0.00002	0.00000	-0.00045	0.00	0.00000	0.00000	-0.00001	0.00000	0.00004	0.00000	0.00	-0.00001	0.00000	0.00000	0.00000
x_3	0.00	0.00002	0.00000	0.00000	0.00000	0.00000	0.00	-0.00004	0.00000	0.00002	0.00000	-0.00004	0.00001	0.00	0.00001	0.00000	0.00001	0.00000
x_4	-0.09	-0.09019	0.00000	-0.09012	0.00000	-0.09066	-0.09	-0.08986	0.00000	-0.09069	0.00000	-0.09062	0.00007	-0.09	-0.08999	0.00000	-0.09000	0.00000
x_5	0.00	0.00000	0.00000	0.00000	0.00000	0.00009	0.00	-0.00001	0.00000	-0.00004	0.00000	0.00013	0.00002	0.00	0.00000	0.00000	-0.00003	0.00000
x_6	0.10	0.10012	0.00000	0.09953	0.00000	0.09778	0.10	0.10013	0.00000	0.09976	0.00000	0.09941	0.00008	0.10	0.10033	0.00000	0.10091	0.00000
w_1	0.00	-0.00001	0.00000	0.00005	0.00000	-0.00001	0.00	-0.00001	0.00000	0.00003	0.00000	0.00006	0.00001	0.00	0.00008	0.00000	-0.00002	0.00000
w_2	-0.10	-0.09922	0.00001	-0.09932	0.00001	-0.09967	-0.10	-0.09863	0.00001	-0.09944	0.00001	-0.09917	0.00001	-0.10	-0.09915	0.00001	-0.09725	0.00001
$N=2000$																		
y_{t-1}	0.95	0.94980	0.00000	0.94975	0.00000	0.00010	0.50	0.50463	0.00004	0.50147	0.00008	0.49932	0.00016	0.30	0.30293	0.00004	0.30625	0.00009
x_1	0.07	0.06970	0.00000	0.07034	0.00000	0.00004	0.07	0.07015	0.00000	0.07017	0.00000	0.07059	0.00004	0.07	0.07021	0.00000	0.07016	0.00000
x_2	0.00	-0.00001	0.00000	0.00000	0.00000	0.00000	0.00	0.00000	0.00000	0.00002	0.00000	0.00001	0.00000	0.00	0.00000	0.00000	-0.00001	0.00000
x_3	0.00	0.00002	0.00000	-0.00001	0.00000	0.00000	0.00	-0.00002	0.00000	0.00000	0.00000	0.00008	0.00000	0.00	0.00000	0.00000	0.00000	0.00000
x_4	-0.09	-0.09007	0.00000	-0.09016	0.00000	-0.08839	-0.09	-0.08986	0.00000	-0.09062	0.00000	-0.09010	0.00005	-0.09	-0.08987	0.00000	-0.09048	0.00000
x_5	0.00	0.00000	0.00000	0.00002	0.00000	0.00020	0.00	0.00000	0.00000	-0.00002	0.00000	0.00010	0.00001	0.00	0.00001	0.00000	-0.00003	0.00000
x_6	0.10	0.10010	0.00000	0.09897	0.00000	0.10001	0.10	0.10012	0.00000	0.09956	0.00000	0.09884	0.00003	0.10	0.10045	0.00000	0.10070	0.00000
w_1	0.00	-0.00002	0.00000	-0.00003	0.00000	0.00003	0.00	-0.00001	0.00000	0.00002	0.00000	0.00010	0.00001	0.00	0.00002	0.00000	-0.00002	0.00000
w_2	-0.10	-0.09988	0.00000	-0.09972	0.00000	-0.09993	-0.10	-0.09806	0.00000	-0.09919	0.00000	-0.10029	0.00001	-0.10	-0.09913	0.00000	-0.09816	0.00000

Notes:

1. For the idiosyncratic error term, $\eta_t \sim N(0, \sigma_\eta^2)$ where $\sigma_\eta = 0.10$.
2. The error term is normally distributed $v_{it} \sim N(0, \sigma_v^2)$.

Table 6. Posterior Probability of the True Model
Summary statistics using LIBMA estimation for various n , α , and σ_v

<i>Sample</i>	α	0.95			0.50			0.30		
		σ_v			σ_v			σ_v		
		0.05	0.10	0.50	0.05	0.10	0.50	0.05	0.10	0.50
<i>N=200</i>										
Mean		0.344	0.318	0.240	0.349	0.328	0.262	0.332	0.319	0.252
Variance		0.017	0.015	0.017	0.014	0.021	0.017	0.017	0.018	0.019
Q1		0.273	0.223	0.127	0.284	0.231	0.145	0.229	0.212	0.146
Median		0.379	0.321	0.244	0.370	0.355	0.278	0.360	0.349	0.253
Q3		0.454	0.414	0.342	0.449	0.453	0.356	0.435	0.433	0.366
<i>N=500</i>										
Mean		0.476	0.470	0.437	0.474	0.460	0.479	0.478	0.501	0.470
Variance		0.025	0.020	0.026	0.019	0.027	0.018	0.022	0.017	0.021
Q1		0.410	0.397	0.378	0.393	0.373	0.386	0.388	0.454	0.402
Median		0.523	0.499	0.463	0.515	0.518	0.534	0.528	0.538	0.510
Q3		0.604	0.594	0.567	0.585	0.588	0.580	0.585	0.603	0.584
<i>N=1000</i>										
Mean		0.532	0.538	0.549	0.549	0.564	0.553	0.564	0.578	0.567
Variance		0.031	0.031	0.028	0.026	0.023	0.022	0.027	0.028	0.020
Q1		0.406	0.451	0.509	0.493	0.492	0.497	0.516	0.525	0.518
Median		0.603	0.578	0.595	0.597	0.605	0.587	0.624	0.642	0.610
Q3		0.664	0.675	0.668	0.662	0.672	0.657	0.674	0.692	0.669
<i>N=2000</i>										
Mean		0.598	0.633	0.606	0.626	0.657	0.667	0.651	0.668	0.660
Variance		0.039	0.019	0.028	0.034	0.021	0.016	0.020	0.014	0.018
Q1		0.506	0.580	0.513	0.562	0.624	0.627	0.598	0.630	0.638
Median		0.669	0.669	0.657	0.699	0.710	0.717	0.708	0.706	0.699
Q3		0.742	0.734	0.737	0.756	0.753	0.752	0.742	0.751	0.746

Notes:

1. The error terms are constructed using discrete distributions (see section IV.A.).

Table 7. Posterior Probability Ratio: True Model/Best Among the Other Models
Summary statistics using LIBMA estimation for various n , α , and σ_v

<i>Sample</i>	α	0.95			0.50			0.30		
		σ_v			σ_v			σ_v		
		0.05	0.10	0.50	0.05	0.10	0.50	0.05	0.10	0.50
<i>N=200</i>										
Mean		2.282	2.040	1.659	2.323	2.242	1.798	2.175	2.080	1.820
Variance		1.981	2.000	1.606	1.726	2.321	1.629	1.803	1.932	1.822
Q1		1.230	0.867	0.690	1.304	0.949	0.730	1.047	0.873	0.707
Median		2.111	1.823	1.353	2.196	1.908	1.522	2.152	2.047	1.461
Q3		3.502	3.050	2.417	3.348	3.466	2.475	3.213	3.064	2.690
<i>N=500</i>										
Mean		3.738	3.567	3.175	3.563	3.596	3.681	3.704	3.882	3.504
Variance		5.286	4.471	4.485	4.499	5.716	4.507	4.530	4.605	4.756
Q1		1.784	1.809	1.526	2.031	1.564	1.797	1.651	2.021	1.572
Median		3.809	3.196	2.959	3.128	3.430	3.899	3.860	3.682	3.570
Q3		5.677	5.719	4.581	5.223	5.279	5.189	5.422	5.816	5.368
<i>N=1000</i>										
Mean		4.482	4.651	4.645	4.578	5.034	4.693	4.995	5.686	4.969
Variance		9.738	10.456	8.166	8.924	9.303	9.957	8.401	11.604	8.523
Q1		1.634	1.813	2.436	2.232	2.226	2.468	2.657	2.674	2.826
Median		4.213	3.746	4.760	3.775	4.945	4.029	5.106	5.878	4.703
Q3		6.376	7.137	6.650	6.789	7.405	6.868	7.220	8.211	7.094
<i>N=2000</i>										
Mean		6.061	5.997	5.740	6.569	7.008	7.180	6.699	7.293	6.669
Variance		18.867	14.406	16.746	19.530	15.723	15.063	14.792	16.717	13.738
Q1		2.663	2.953	1.912	3.268	4.008	4.345	3.288	3.937	3.740
Median		5.228	5.302	5.160	6.278	6.977	7.028	6.701	6.917	6.427
Q3		9.679	8.340	8.646	9.696	9.771	10.033	9.571	10.203	9.015

Notes:

1. The error terms are constructed using discrete distributions (see section IV.A.).

Table 8. Probability of Retrieving the True Model
Summary statistics using LIBMA estimation for various n , α , and σ_v

<i>Sample</i>	α	0.95			0.50			0.30		
		σ_v			σ_v			σ_v		
		0.05	0.10	0.50	0.05	0.10	0.50	0.05	0.10	0.50
<hr/>										
<i>N=200</i>										
% Correct		78	70	65	82	72	71	76	70	66
<i>N=500</i>										
% Correct		87	87	81	90	84	88	85	91	83
<i>N=1000</i>										
% Correct		82	88	89	88	91	87	87	90	92
<i>N=2000</i>										
% Correct		88	94	89	89	92	92	96	97	96
<hr/>										

Notes:

1. The error terms are constructed using discrete distributions (see section IV.A.).

Table 9. Model Recovery: Medians and Variances of Posterior Inclusion Probability for Each Variable
True model vs BMA posterior inclusion probability for various n , α , and σ_v

Sample	α σ_v	0.95								0.50				0.30							
		0.05		0.10		0.50		0.05		0.10		0.50		0.05		0.10		0.50			
		Model	Median	Variance	Median	Variance	Median	Variance	Median	Variance	Median	Variance	Median	Variance	Median	Variance	Median	Variance	Median	Variance	
N=200																					
y_{t-1}	1	1.00000	0.00645	0.99999	0.00020	1.00000	0.00000	1.00000	0.00000	1.00000	0.00276	1.00000	0.00000	1.00000	0.00000	0.99999	0.00298	0.99974	0.00428		
x_1	1	1.00000	0.00464	0.99999	0.00003	0.93729	0.04238	1.00000	0.00000	1.00000	0.00000	0.95587	0.04862	1.00000	0.00000	1.00000	0.00000	0.93931	0.04225		
x_2	0	0.17454	0.02173	0.17474	0.01661	0.17118	0.01707	0.18556	0.00762	0.17447	0.02088	0.18673	0.02439	0.17292	0.01695	0.17668	0.01679	0.16842	0.01449		
x_3	0	0.17667	0.02116	0.19481	0.01840	0.18169	0.02376	0.19783	0.01477	0.17500	0.02482	0.19002	0.01603	0.18012	0.01559	0.17240	0.01023	0.17422	0.02975		
x_4	1	1.00000	0.00001	1.00000	0.00005	0.99655	0.02350	1.00000	0.00000	1.00000	0.00000	0.99874	0.01165	1.00000	0.00000	1.00000	0.00000	0.99580	0.00891		
x_5	0	0.17948	0.01703	0.19036	0.02538	0.17095	0.00980	0.17557	0.01284	0.17196	0.01868	0.18045	0.00763	0.18280	0.02120	0.19580	0.03906	0.19070	0.01876		
x_6	1	1.00000	0.00264	1.00000	0.00001	0.99933	0.00683	1.00000	0.00000	1.00000	0.00000	0.99965	0.00122	1.00000	0.00000	1.00000	0.00000	0.99897	0.00719		
w_1	0	0.16520	0.01047	0.17782	0.02025	0.21088	0.03771	0.18629	0.02511	0.19334	0.03158	0.19282	0.03017	0.19512	0.02774	0.17923	0.02152	0.18846	0.03213		
w_2	1	1.00000	0.00004	0.99999	0.00008	0.99896	0.01245	1.00000	0.00000	1.00000	0.00000	0.99871	0.01455	1.00000	0.00000	1.00000	0.00002	0.99905	0.01706		
N=500																					
y_{t-1}	1	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000		
x_1	1	1.00000	0.00000	1.00000	0.00000	0.99998	0.00070	1.00000	0.00000	1.00000	0.00000	0.99998	0.00009	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000		
x_2	0	0.12089	0.01624	0.12964	0.01162	0.12260	0.01910	0.12530	0.01152	0.12395	0.01814	0.11977	0.01275	0.12902	0.01562	0.11685	0.01298	0.11901	0.01519		
x_3	0	0.11402	0.00929	0.13206	0.01191	0.13164	0.01376	0.12678	0.01312	0.11890	0.02294	0.12262	0.01901	0.11613	0.00696	0.11594	0.00213	0.12111	0.02191		
x_4	1	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000		
x_5	0	0.11929	0.02669	0.11284	0.02250	0.12951	0.03295	0.12776	0.02185	0.12074	0.01966	0.13283	0.01079	0.12515	0.01542	0.11329	0.01032	0.11822	0.01197		
x_6	1	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000		
w_1	0	0.11260	0.01555	0.12982	0.01207	0.13208	0.02213	0.12110	0.00725	0.12079	0.01688	0.11993	0.01018	0.11700	0.02356	0.12405	0.01790	0.12476	0.01897		
w_2	1	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000		
N=1000																					
y_{t-1}	1	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000		
x_1	1	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000		
x_2	0	0.08629	0.01338	0.08669	0.01975	0.09677	0.01757	0.09342	0.01974	0.09442	0.00916	0.09734	0.00960	0.08971	0.00402	0.08278	0.00666	0.09309	0.01893		
x_3	0	0.09391	0.01813	0.08700	0.01796	0.09776	0.01077	0.09500	0.02258	0.10320	0.02065	0.08946	0.01944	0.08979	0.01938	0.08941	0.01775	0.09020	0.01212		
x_4	1	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000		
x_5	0	0.09265	0.02334	0.09357	0.01437	0.08951	0.01987	0.08936	0.01272	0.08673	0.00974	0.09722	0.01832	0.09423	0.01871	0.08965	0.02386	0.09219	0.00721		
x_6	1	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000		
w_1	0	0.08340	0.02110	0.09592	0.01865	0.10312	0.01823	0.08615	0.00752	0.08598	0.00875	0.09037	0.00482	0.08845	0.01810	0.08504	0.01168	0.08700	0.00789		
w_2	1	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000		
N=2000																					
y_{t-1}	1	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000		
x_1	1	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000		
x_2	0	0.06555	0.01849	0.06911	0.00717	0.07534	0.01617	0.06574	0.01595	0.06419	0.01371	0.07118	0.00780	0.06999	0.01023	0.06780	0.00451	0.06292	0.00453		
x_3	0	0.07132	0.01704	0.07693	0.00665	0.06512	0.00897	0.06166	0.01239	0.07266	0.00357	0.06714	0.00657	0.06812	0.00899	0.06538	0.00607	0.07263	0.02124		
x_4	1	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000		
x_5	0	0.07611	0.02266	0.06596	0.00551	0.06860	0.01023	0.07207	0.01844	0.06466	0.01515	0.06562	0.00570	0.06948	0.00422	0.06723	0.00586	0.06441	0.00261		
x_6	1	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000		
w_1	0	0.06935	0.02144	0.07806	0.01694	0.07013	0.02338	0.06401	0.01526	0.06688	0.00465	0.06322	0.00984	0.06317	0.01193	0.06128	0.00944	0.06257	0.00458		
w_2	1	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000		

Notes:

1. The error terms are constructed using discrete distributions (see section IV.A.).

Table 10. Model Recovery: Medians and Variances of Estimated Parameter Values
True model vs BMA coefficients' estimated values for various n , α , and σ_v

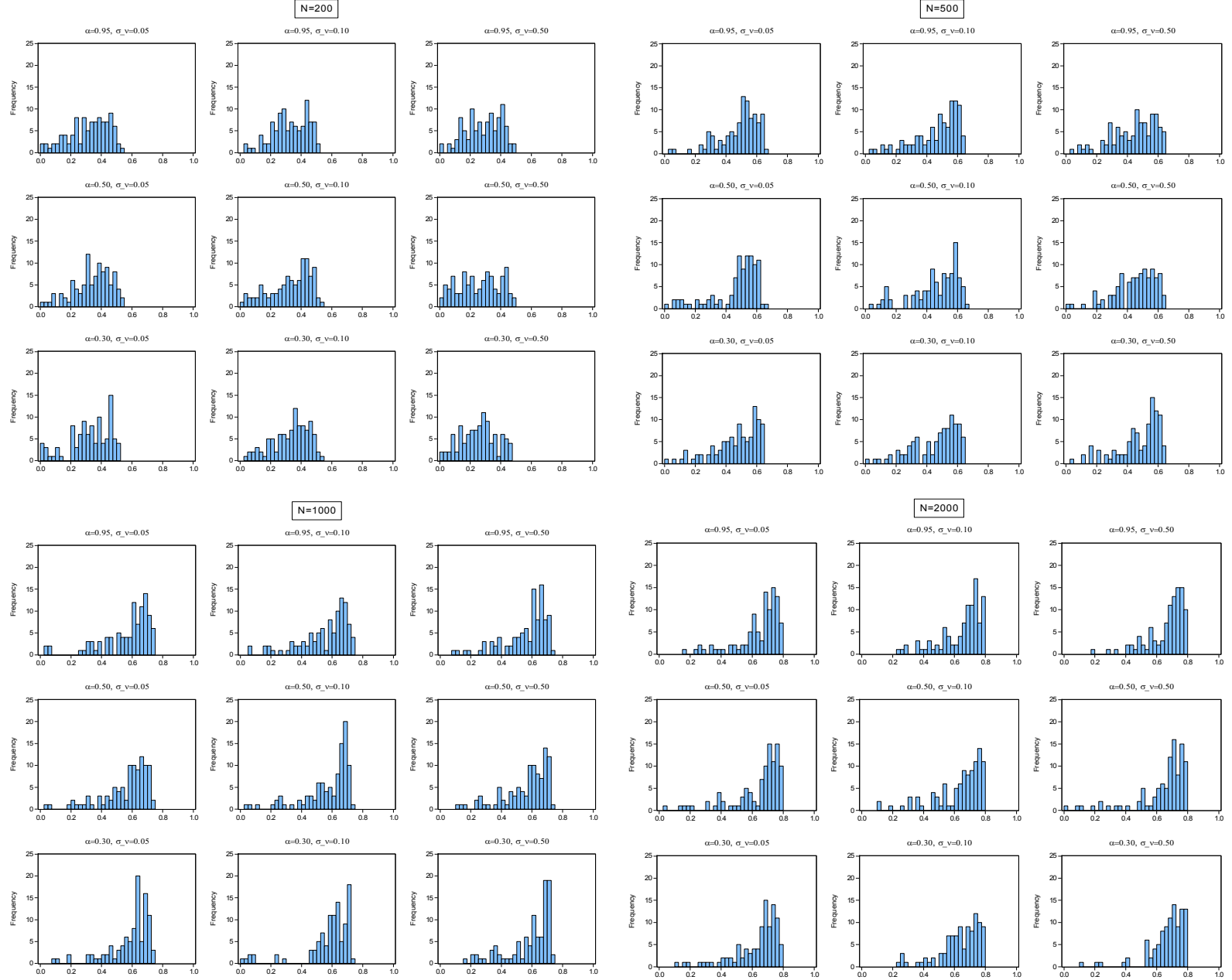
Sample	α σ_v	0.95								0.50								0.30							
		True		0.05000		0.10000		0.50000		True		0.05000		0.10000		0.50000		True		0.05000		0.10000		0.50000	
		Value	Median	Variance	Median	Variance	Median	Variance	Value	Median	Variance	Median	Variance	Median	Variance	Value	Median	Variance	Median	Variance	Median	Variance	Median	Variance	
N=200																									
y _{t-1}	0.95	0.95000	0.00586	0.94907	0.00024	0.95641	0.00136	0.50	0.51392	0.00056	0.50828	0.00227	0.53691	0.00200	0.30	0.31301	0.00058	0.32632	0.00151	0.33918	0.00299				
x ₁	0.07	0.07003	0.00004	0.07037	0.00002	0.06381	0.00081	0.07	0.07016	0.00001	0.07107	0.00002	0.06747	0.00074	0.07	0.06952	0.00001	0.07038	0.00002	0.06403	0.00066				
x ₂	0.00	0.00004	0.00000	0.00001	0.00000	-0.00081	0.00009	0.00	-0.00003	0.00000	0.00003	0.00001	0.00066	0.00015	0.00	0.00007	0.00000	-0.00010	0.00000	0.00009	0.00008				
x ₃	0.00	0.00001	0.00000	0.00002	0.00000	-0.00042	0.00016	0.00	0.00000	0.00000	-0.00008	0.00001	0.00000	0.00009	0.00	0.00000	0.00000	0.00007	0.00000	-0.00057	0.00015				
x ₄	-0.09	-0.09019	0.00003	-0.09031	0.00001	-0.08839	0.00079	-0.09	-0.08976	0.00001	-0.08976	0.00002	-0.09138	0.00070	-0.09	-0.09083	0.00001	-0.09052	0.00002	-0.08948	0.00055				
x ₅	0.00	0.00008	0.00000	0.00002	0.00001	-0.00009	0.00005	0.00	0.00001	0.00000	-0.00003	0.00000	0.00060	0.00004	0.00	0.00003	0.00000	-0.00015	0.00001	-0.00016	0.00010				
x ₆	0.10	0.09953	0.00003	0.09898	0.00002	0.09978	0.00068	0.10	0.10044	0.00001	0.10002	0.00002	0.10080	0.00045	0.10	0.10020	0.00001	0.09973	0.00002	0.09942	0.00060				
w ₁	0.00	0.00028	0.00001	0.00059	0.00003	0.00132	0.00040	0.00	0.00022	0.00001	0.00036	0.00002	-0.00073	0.00024	0.00	0.00026	0.00001	0.00016	0.00001	0.00007	0.00032				
w ₂	-0.10	-0.09635	0.00127	-0.09712	0.00006	-0.10031	0.00041	-0.10	-0.09567	0.00003	-0.09692	0.00006	-0.09428	0.00028	-0.10	-0.09809	0.00004	-0.09405	0.00004	-0.09610	0.00035				
N=500																									
y _{t-1}	0.95	0.95000	0.00001	0.95010	0.00001	0.95674	0.00048	0.50	0.50875	0.00023	0.50203	0.00065	0.51779	0.00070	0.30	0.30944	0.00020	0.31437	0.00044	0.32502	0.00072				
x ₁	0.07	0.07012	0.00000	0.06996	0.00001	0.06880	0.00018	0.07	0.06966	0.00000	0.07034	0.00001	0.06715	0.00017	0.07	0.07011	0.00000	0.06975	0.00001	0.07337	0.00014				
x ₂	0.00	-0.00001	0.00000	0.00004	0.00000	-0.00031	0.00004	0.00	-0.00001	0.00000	0.00002	0.00000	-0.00017	0.00002	0.00	0.00003	0.00000	-0.00007	0.00000	-0.00015	0.00003				
x ₃	0.00	-0.00004	0.00000	0.00009	0.00000	-0.00018	0.00002	0.00	0.00002	0.00000	-0.00008	0.00000	-0.00009	0.00003	0.00	0.00001	0.00000	-0.00003	0.00000	-0.00060	0.00003				
x ₄	-0.09	-0.09016	0.00000	-0.09064	0.00001	-0.09000	0.00014	-0.09	-0.09021	0.00000	-0.09006	0.00001	-0.09240	0.00017	-0.09	-0.09063	0.00000	-0.09086	0.00001	-0.08872	0.00017				
x ₅	0.00	0.00003	0.00000	0.00001	0.00000	0.00015	0.00007	0.00	-0.00004	0.00000	0.00008	0.00000	0.00003	0.00002	0.00	0.00002	0.00000	-0.00005	0.00000	-0.00013	0.00002				
x ₆	0.10	0.10020	0.00000	0.09925	0.00001	0.10256	0.00016	0.10	0.10023	0.00000	0.09945	0.00001	0.10119	0.00015	0.10	0.09986	0.00000	0.09960	0.00001	0.09951	0.00016				
w ₁	0.00	0.00005	0.00000	0.00019	0.00000	0.00058	0.00004	0.00	0.00008	0.00000	0.00008	0.00000	-0.00012	0.00001	0.00	0.00017	0.00000	0.00004	0.00000	0.00007	0.00003				
w ₂	-0.10	-0.09861	0.00001	-0.09897	0.00002	-0.10023	0.00004	-0.10	-0.09774	0.00001	-0.09811	0.00002	-0.09741	0.00002	-0.10	-0.09876	0.00002	-0.09795	0.00002	-0.09770	0.00003				
N=1000																									
y _{t-1}	0.95	0.95008	0.00000	0.94962	0.00001	0.95328	0.00024	0.50	0.50535	0.00012	0.50579	0.00023	0.51277	0.00026	0.30	0.30580	0.00010	0.30667	0.00022	0.31089	0.00027				
x ₁	0.07	0.07005	0.00000	0.07035	0.00000	0.06885	0.00006	0.07	0.07000	0.00000	0.07026	0.00000	0.07080	0.00009	0.07	0.06988	0.00000	0.06904	0.00000	0.07201	0.00009				
x ₂	0.00	0.00002	0.00000	0.00002	0.00000	-0.00019	0.00001	0.00	0.00002	0.00000	0.00004	0.00000	-0.00002	0.00001	0.00	0.00002	0.00000	-0.00001	0.00000	-0.00015	0.00002				
x ₃	0.00	-0.00002	0.00000	0.00007	0.00000	-0.00021	0.00001	0.00	0.00002	0.00000	-0.00009	0.00000	-0.00002	0.00002	0.00	0.00000	0.00000	0.00004	0.00000	-0.00035	0.00001				
x ₄	-0.09	-0.09038	0.00000	-0.09047	0.00000	-0.08988	0.00007	-0.09	-0.08995	0.00000	-0.08961	0.00000	-0.09250	0.00006	-0.09	-0.09011	0.00000	-0.09005	0.00000	-0.08765	0.00008				
x ₅	0.00	0.00004	0.00000	0.00000	0.00000	-0.00007	0.00002	0.00	-0.00003	0.00000	0.00000	0.00000	-0.00016	0.00001	0.00	0.00002	0.00000	-0.00002	0.00000	-0.00011	0.00001				
x ₆	0.10	0.10008	0.00000	0.09951	0.00000	0.10166	0.00007	0.10	0.10034	0.00000	0.10010	0.00000	0.10064	0.00008	0.10	0.10022	0.00000	0.09948	0.00000	0.09632	0.00008				
w ₁	0.00	0.00004	0.00000	0.00021	0.00000	0.00058	0.00002	0.00	0.00001	0.00000	-0.00002	0.00000	-0.00019	0.00000	0.00	0.00001	0.00000	-0.00002	0.00000	0.00000	0.00001				
w ₂	-0.10	-0.09959	0.00001	-0.09913	0.00001	-0.10109	0.00002	-0.10	-0.09808	0.00001	-0.09848	0.00001	-0.09808	0.00001	-0.10	-0.09876	0.00001	-0.09915	0.00001	-0.09837	0.00001				
N=2000																									
y _{t-1}	0.95	0.94994	0.00000	0.94972	0.00000	0.95469	0.00011	0.50	0.50489	0.00004	0.49944	0.00011	0.51104	0.00011	0.30	0.30273	0.00006	0.30012	0.00010	0.31344	0.00015				
x ₁	0.07	0.07026	0.00000	0.07036	0.00000	0.06932	0.00004	0.07	0.06967	0.00000	0.07018	0.00000	0.07050	0.00003	0.07	0.06978	0.00000	0.06914	0.00000	0.07292	0.00004				
x ₂	0.00	0.00000	0.00000	0.00000	0.00000	-0.00037	0.00001	0.00	0.00000	0.00000	0.00002	0.00000	-0.00002	0.00000	0.00	0.00001	0.00000	-0.00002	0.00000	-0.00004	0.00000				
x ₃	0.00	-0.00003	0.00000	0.00006	0.00000	-0.00013	0.00000	0.00	0.00002	0.00000	-0.00005	0.00000	-0.00004	0.00000	0.00	-0.00001	0.00000	0.00003	0.00000	-0.00031	0.00001				
x ₄	-0.09	-0.09020	0.00000	-0.09000	0.00000	-0.09002	0.00003	-0.09	-0.09004	0.00000	-0.08991	0.00000	-0.09076	0.00003	-0.09	-0.09018	0.00000	-0.09040	0.00000	-0.08852	0.00004				
x ₅	0.00	0.00005	0.00000	0.00000	0.00000	-0.00006	0.00000	0.00	-0.00002	0.00000	0.00001	0.00000	-0.00005	0.00000	0.00	0.00002	0.00000	-0.00002	0.00000	-0.00007	0.00000				
x ₆	0.10	0.10012	0.00000	0.09952	0.00000	0.10034	0.00004	0.10	0.10052	0.00000	0.09990	0.00000	0.09883	0.00003	0.10	0.10024	0.00000	0.09974	0.00000	0.09679	0.00003				
w ₁	0.00	0.00001	0.00000	0.00013	0.00000	0.00022	0.00001	0.00	-0.00002	0.00000	-0.00001	0.00000	-0.00006	0.00000	0.00	0.00003	0.00000	-0.00002	0.00000	-0.00003	0.00000				
w ₂	-0.10	-0.09988	0.00000	-0.10014	0.00000	-0.10043	0.00001	-0.10	-0.09849	0.00000	-0.09920	0.00000	-0.09881	0.00000	-0.10	-0.09941	0.00000	-0.09969	0.00000	-0.09884	0.00000				

Notes:

1. The error terms are constructed using discrete distributions (see section IV.A.).

Appendix A

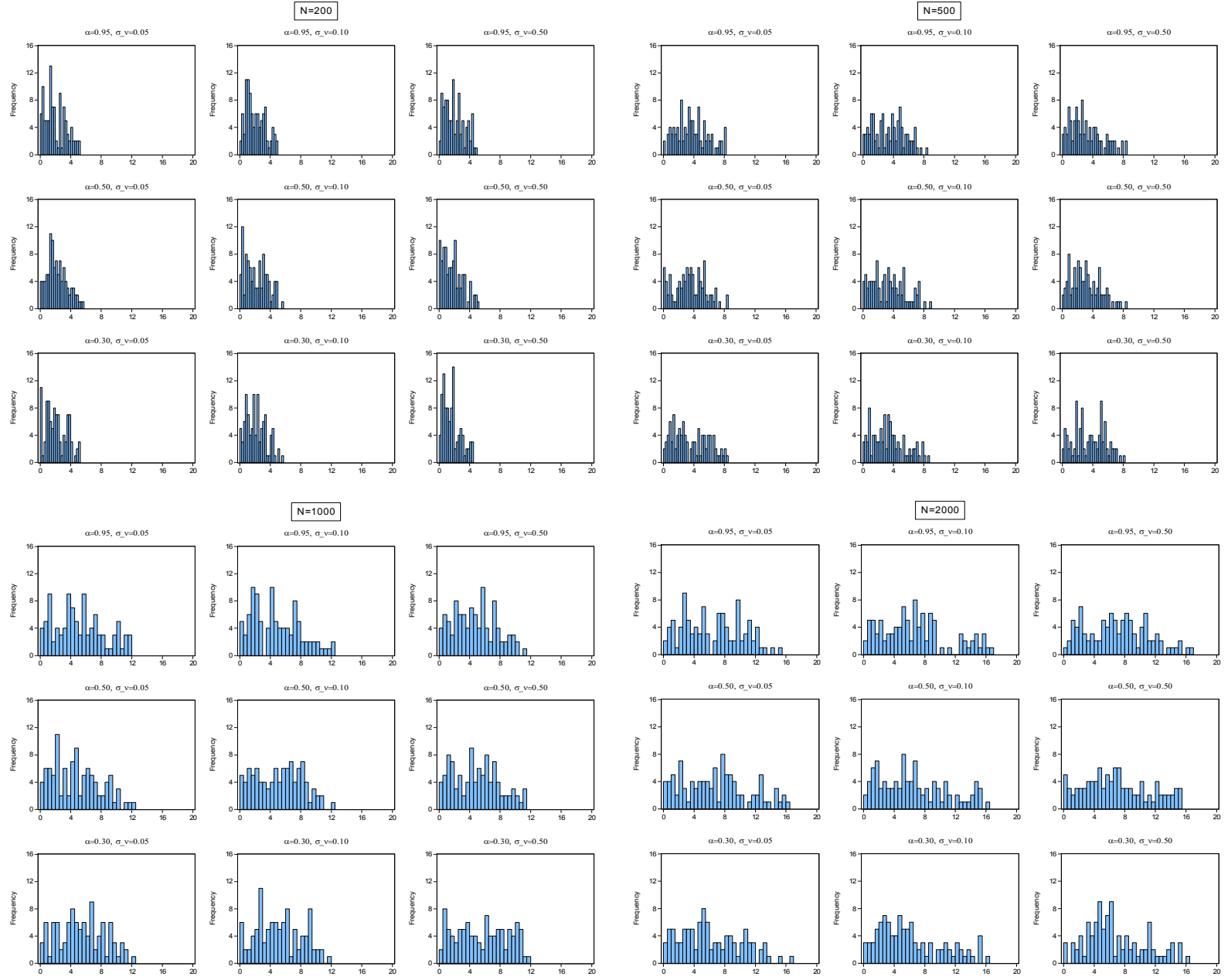
Figure 1: Posterior Densities for the Probabilities in Table 1



Notes:

1. For the idiosyncratic error term, $\eta_i \sim N(0, \sigma_\eta^2)$ where $\sigma_\eta = 0.10$.
2. The error term is normally distributed $\psi_i \sim N(0, \sigma_\psi^2)$.

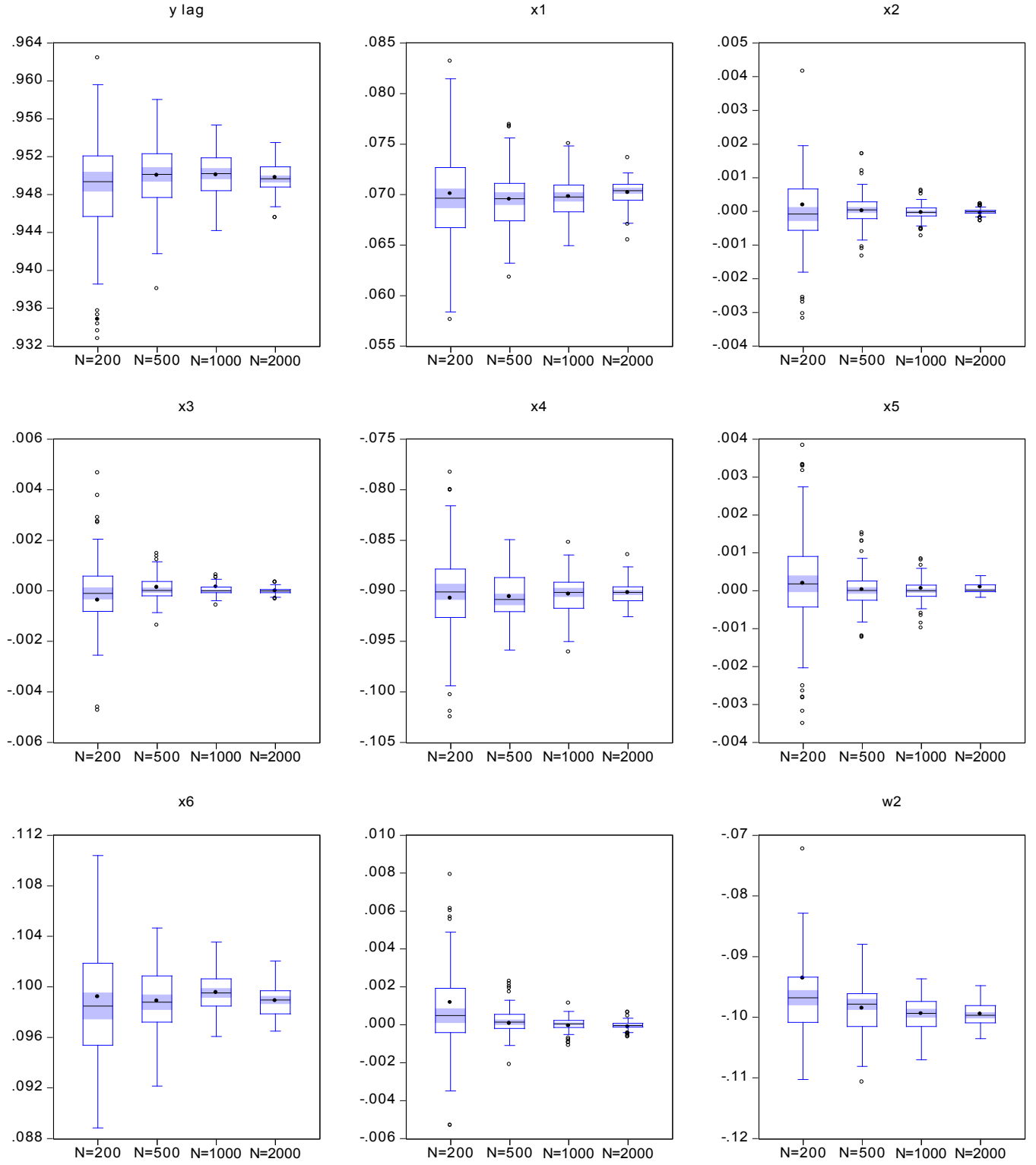
Figure 2: Posterior Densities for the Probabilities in Table 2



Notes:

1. For the idiosyncratic error term, $\eta_i \sim N(0, \sigma_\eta^2)$ where $\sigma_\eta = 0.10$.
2. The error term is normally distributed $v_i \sim N(0, \sigma_v^2)$.

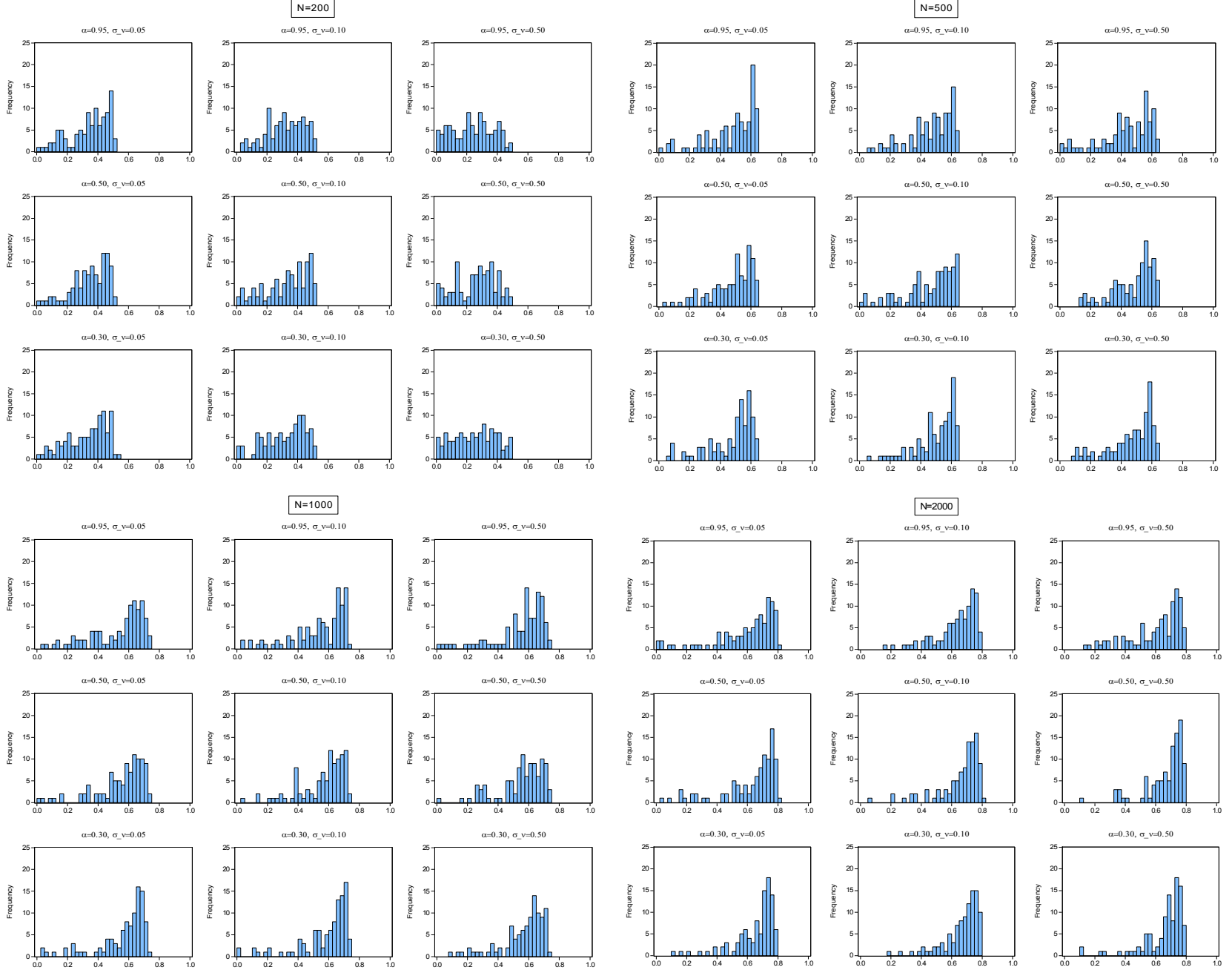
Figure 3: Box Plots for Parameters in Table 5
BMA coefficients' estimated values for $\alpha=0.95$, $\sigma_v=0.1$ and various N



Notes:

1. For the idiosyncratic error term, $\eta_i \sim N(0, \sigma_\eta^2)$ where $\sigma_\eta = 0.10$.
2. The error term is normally distributed $v_{it} \sim N(0, \sigma_v^2)$.

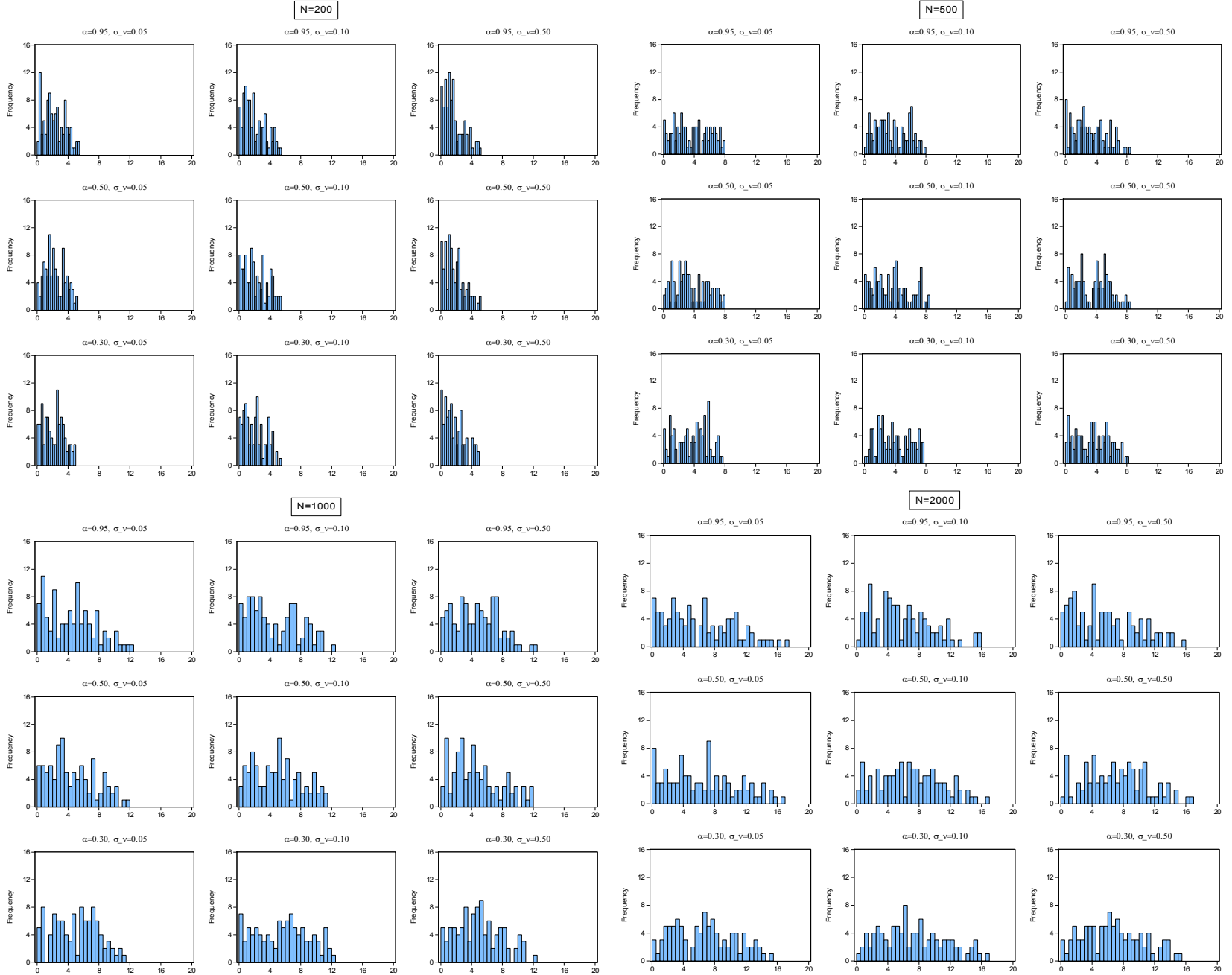
Figure 3: Posterior Densities for the Probabilities in Table 6



Notes:

1. The error terms are constructed using discrete distributions (see section IV.A.).

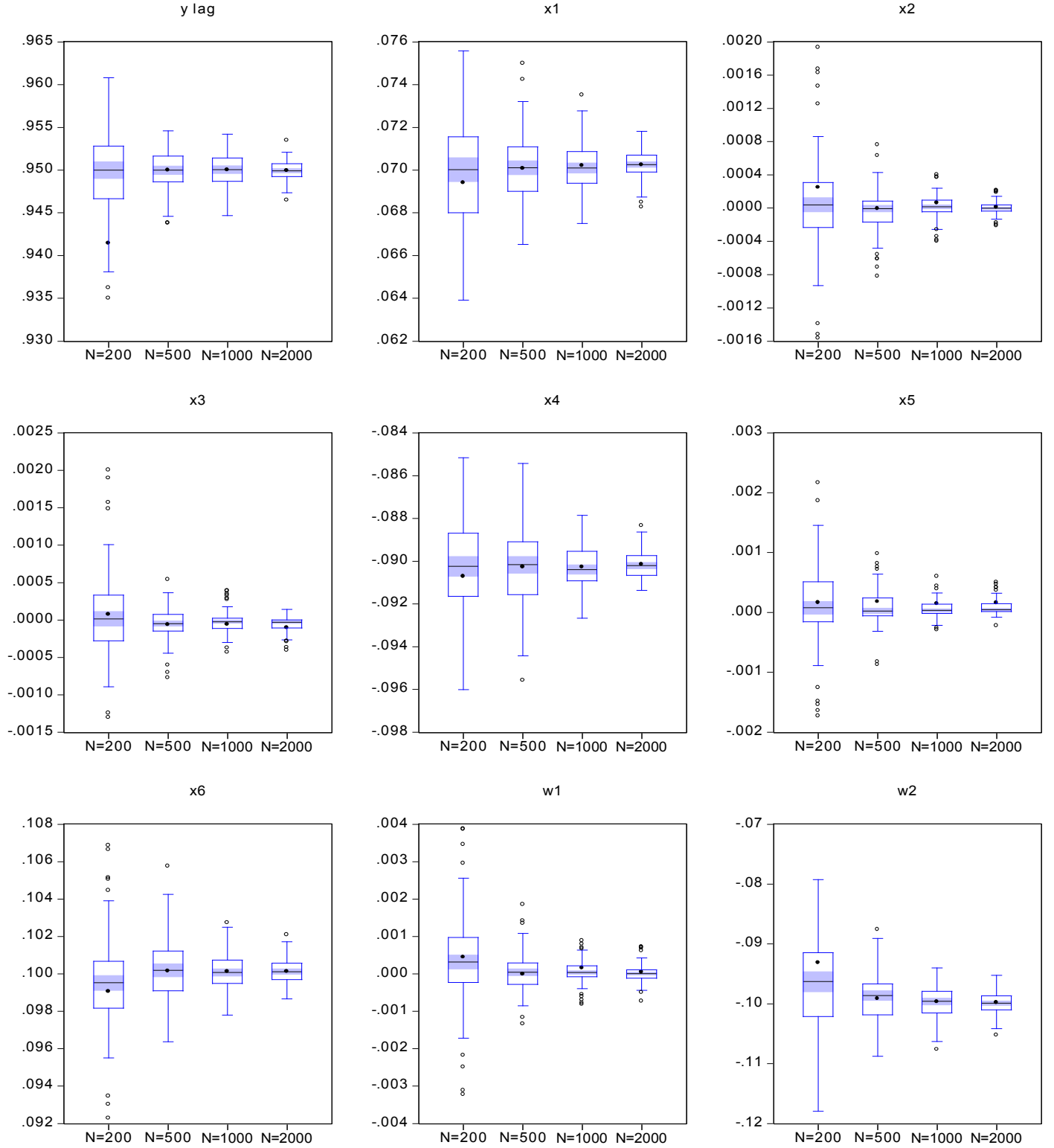
Figure 4: Posterior Densities for the Probabilities in Table 7



Notes:

1. The error terms are constructed using discrete distributions (see section IV.A.).

Figure 6: Box Plots for Parameters in Table 10
BMA coefficients' estimated values for $\alpha=0.95$, $\sigma_v=0.1$ and various N



Notes:

1. The error terms are constructed using discrete distributions (see section IV.A.).