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

Huigang Chen
International Monetary Fund

Alin Mirestean
International Monetary Fund

Charalambos G. Tsangarides
International Monetary Fund

June 30, 2009
Preliminary

Abstract

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.

---

*We thank Michael Binder, Steve Durlauf, Rex Ghosh, Eduardo Ley, Enrique Moral-Benito, Chris Papageorgiou, and Roberto Samaniego for helpful comments and suggestions. This paper should not be reported as representing the views of the IMF. The views expressed are those of the authors and do not necessarily reflect the views of the IMF or IMF policy. An earlier version of this paper appeared as IMF Working Paper No. 09/74. Emails: hchen@imf.org, amirestean@imf.org, ctsangarides@imf.org.

1 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 trade-offs 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.[1] 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.[2] 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.[3] Moreover, to the best of our knowledge, none of the models allow for the inclusion of endogenous variables.

In this paper we propose a methodology for dealing with model uncertainty in the context of a panel data model with short time periods where the lagged dependent variable as well as 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). Specifically, we propose a method for constructing the model likelihoods and posteriors based only on information elicited from moment conditions. We evaluate 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.

---

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

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

[3] 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.

2 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).

2.1 Model Selection and Hypothesis Testing

Consider the standard linear regression model

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

where Y is the variable of interest, $\breve{Z}$ 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 $U = \{1, 2, ..., j, j+1, ..., k\}$. Let 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, $M_j \subset U$, let $C_{M_j} = \left\{c_{mn,M_j}\right\}_{m,n=1}^{k}$ be a $k \times k$ diagonal choice matrix such that its diagonal will have 1's if the corresponding variable is included in the model and 0's otherwise. Hence $c_{ii,M_j} = 1\left\{i \in M_j\right\}$. Therefore, for a given model M_j, $\breve{Z} = ZC_{M_j}$ and model (1) can be now written more generally as

$$Y = ZC_{M_j}\theta + u \tag{2}$$

where $\theta = \begin{pmatrix} \theta_1 & \theta_2 & ... & \theta_k \end{pmatrix}'$ is the set of parameters to be estimated.

Given the universe of k possible explanatory variables, a set of $K = 2^k$ models $\mathcal{M} = (M_1, ..., M_K)$ are under consideration. In the spirit of Bayesian inference, one can specify priors $p(\theta|M_j)$ 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 $\mathcal{M} = (M_1, ..., M_K)$ that actually generated the data. Let $D = \begin{pmatrix} Y & Z \end{pmatrix}$ 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

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

where

$$p(D|M_j) = \int p(D|\theta_j, M_j)p(\theta_j|M_j)d\theta_j \tag{4}$$

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 $\frac{p(M_j|D)}{p(M_i|D)} = \frac{p(D|M_j)}{p(D|M_i)} \cdot \frac{p(M_j)}{p(M_i)}$. Essentially, the data updates the prior odds ratio $\frac{p(M_j)}{p(M_i)}$ through the Bayes factor $\frac{p(D|M_j)}{p(D|M_i)}$ to measure the extent to which the data support M_j over M_i. When the posterior odds ratio is greater (less) then 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,[4] in which case the posterior odds ratio is equal to the Bayes factor B_{ji}.

2.2 Bayesian Model Averaging

A natural strategy for model selection is to chose 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 tends to perform better than other variable selection methods in terms of predictive performance.

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

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

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 $(M_1, ..., M_K)$, where the weights are the posterior model probabilities $p(M_j|D)$. Going back to the linear regression model (2) BMA allows the computation of the inclusion probability for every possible explanatory variable.

$$p(Z_i|D) = \sum_{j=1}^{K} I(Z_i|M_j)p(M_j|D) \tag{6}$$

where

$$I(Z_i|M_j) = \left\{ \begin{array}{ccc} 1 & if & Z_i \in M_j \\ 0 & if & Z_i \notin M_j \end{array} \right\} \tag{7}$$

Using (5) one can compute the posterior mean for parameters θ_l as follows.

$$E(\theta_l|D) = \sum_{j=1}^{K} E(\theta_l|D, M_j)p(M_j|D) \tag{8}$$

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, $p\left(\theta|M_i\right)$.

---

[4] As also in Fernández, Ley and Steel (2001b).

2.3 Choice of Priors

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

$$p\left(D|M_j\right) = \int p\left(D|\theta, M_j\right) p\left(\theta|M_j\right) d\theta. \tag{9}$$

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

$$p\left(D|M_i\right) = \int p\left(D|\theta, \zeta, M_i\right) p\left(\theta, \zeta|M_i\right) d\theta d\zeta. \tag{10}$$

In this case, determining the prior $p\left(\theta, \zeta|M_i\right)$ becomes an important issue.[5]

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 $\left(\theta, \sigma_u^{-2}\right)$ is the Normal-Gamma distribution, which has the benefit of rendering a closed-form posterior.[6] 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 $\frac{\gamma}{\lambda^2}$. 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 prior for σ_u with density $p\left(\sigma_u\right) \propto \sigma_u^{-1}$. 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 Zellner's g-prior with 0 mean

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

which can be motivated by the fact that the correlation of the OLS estimate $\widehat{\theta}$ is proportional to $\left(\tilde{Z}'\tilde{Z}\right)^{-1}\sigma_u^2$.

Another choice of prior for Bayes factors when data is an i.i.d. sequence of observations is the so-called unit information prior, which is the prior used throughout this paper. Suppose we have a parameter estimate $\hat{\theta}$ for model M_l. The prior is a k_l dimensional multivariate normal distribution with mean $\hat{\theta}$ and variance $I(\hat{\theta})^{-1}$. Here $I(\hat{\theta})$ is the expected Fisher information matrix at $\hat{\theta}$ for one observation. It is a $k_l \times k_l$ matrix and its (i, j) entry is defined as

$$I_{ij}(\hat{\theta}) = -E_{\hat{\theta}}\left[\frac{\partial^2 p(D_1|\theta, M_l)}{\partial\theta_i\partial\theta_j}\right].$$

We denote one observation from D by D_1. Intuitively, this prior provides roughly the same amount of information that one observation would give on average.

---

[5] 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.

[6] For a more detailed discussion see Kass and Wasserman (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 $p(M_1) = p(M_2) = \ldots = p(M_K) = \frac{1}{K}$. Alternatively, the model priors may reflect the view of the researcher relative to the number of regressors that should be included and may contain a penalty that increases proportional with the number of regressors included in the model. For example, Sala-i-Martin, Doppelhofer and Miller (2004) use a prior model probability structure, initially proposed by Mitchell and Beauchamp (1988), which reflects the researcher's prior about the size of the model. Specifically, the prior probability for model M_j is

$$p\left(M_j\right) = \left(\frac{\overline{k}}{k}\right)^{k_j} \left(1 - \frac{\overline{k}}{k}\right)^{k-k_j} \tag{11}$$

and hence the prior odds ratio is

$$\frac{p(M_j)}{p(M_l)} = \left(\frac{\overline{k}}{k}\right)^{k_j-k_l} \left(1 - \frac{\overline{k}}{k}\right)^{k_l-k_j} \tag{12}$$

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

3 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.

3.1 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 U, consists of the lagged dependent variable, indexed by 1, a set of m exogenous variables, indexed by X, as well as a set of q endogenous variables, indexed by W, such that $\left\{\{1\}, X, W\right\}$ is a partition of U. Therefore, for a given model $M_j \subset U$, (2) becomes

$$\begin{array}{l} y_{it} = \left(\begin{array}{ccc} y_{i,t-1} & x_{it} & w_{it} \end{array}\right) C_{M_j} \left(\begin{array}{ccc} \alpha & \theta_x & \theta_w \end{array}\right)' + u_{it} \\ u_{it} = \eta_i + v_{it} \\ |\alpha| < 1;\ i = 1, 2, ..., N;\ t = 1, 2, ..., T. \end{array} \tag{13}$$

Here y_{it}, x_{it} and 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 $E\left(v_{it}\right) = 0$ and that v_{it}'s are not serially correlated.

x_{it} is a $1 \times m$ vector of exogenous variables while 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, x_{it}, we have

$$E(x_{it}^l v_{is}) = 0, \ \forall i, t, s; \ x_{it}^l \in x_{it} \tag{14}$$

Similarly, for any endogenous variable we have

$$E(w_{it}^l v_{is}) \begin{cases} \neq 0, \ s \leq t \\ = 0, \ \text{otherwise} \end{cases}, \ w_{it}^l \in w_{it}. \tag{15}$$

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

3.2 Estimation and Moment Conditions

A common approach for estimating the model (13) is to use the system GMM framework. This implies constructing the instruments set and moment conditions for the "level equations" (13) and combining them with the moment conditions using the instruments corresponding to the first differences equations. The first differences (FD) equations corresponding to model (13) are given by

$$\begin{array}{l} \Delta y_{it} = \begin{pmatrix} \Delta y_{i,t-1} & \Delta x_{it} & \Delta w_{it} \end{pmatrix} C_{M_j} \begin{pmatrix} \alpha & \theta_x & \theta_w \end{pmatrix}' + \Delta v_{it} \\ |\alpha| < 1; \ i = 1, 2, ..., N; \ \ t = 2, 3, ..., T. \end{array} \tag{16}$$

One assumption required for the FD equations is that the initial value of y, y_{i0}, is predetermined, that is, $E\left(y_{i0} v_{is}\right) = 0$ for $s = 2, 3, ..., T$. Since $y_{i,t-2}$ is not correlated with Δv_{it} we can use it as an instrument. Hence we have $E\left(y_{i,t-2} \Delta v_{it}\right) \neq 0$ for $t = 2, 3, ..., T$. 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

$$E(y_{i,t-s} \Delta v_{it}) = 0, \ t = 2, 3, ..., T; \ s = 2, 3, ..., t; \ \text{for } T \geq 2, \ i = 1, 2, ..., N. \tag{17}$$

Similarly, the exogenous variable x_{it}^l, $x_{it}^l \in x_{it}$ is not correlated with Δv_{it} and therefore we can use it as an instrument.[7] That gives us additional moment conditions

$$E(x_{it}^l \Delta v_{it}) = 0, \ t = 2, 3, ..., T; \ l = 1, ..., m; \ i = 1, 2, ..., N. \tag{18}$$

---

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

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

$$\begin{aligned} E(w^l_{i,t-s}\Delta v_{it}) = 0,\ t = 3, 4, ..., T;\ s = 2, ...t-1; \\ \text{for } T \geq 3,\ l{=}\ 1, 2, ..., q;\ i = 1, ..., N. \end{aligned} \tag{19}$$

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

Table A. Moment Conditions for the First Difference Equation

Variable	Instruments	Moment conditions
$\Delta y_{i,t-1}$	$y_{i,t-2}, ..., y_{i,0}$	$E(y_{i,t-s}\Delta v_{it}) = 0,\ t = 2, 3, ..., T;\ s = 2, 3, ..., t$
Δx^l_{it}	$x^l_{it}, ..., x^l_{i1}$	$E(x^l_{it}\Delta v_{it}) = 0,\ t = 2, 3, ..., T;\ l = 1, 2, ..., m$
Δw^l_{it}	$w^l_{i,t-2}, ..., w^l_{i,1}$	$E(w^l_{i,t-s}\Delta v_{it}) = 0,\ t = 3, 4, ..., T;\ s = 2, 3, ..., t-1;\ l = 1, 2, ..., q$

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 (13), 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

$$E(\Delta y_{i,t-1}u_{it}) = 0,\ t = 2, 3, ..., T. \tag{20}$$

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

$$E(\Delta w^l_{i,t-1}u_{it}) = 0,\ t = 3, 4, ..., T,\ l = 1, 2, ..., q. \tag{21}$$

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

$$E(\Delta x^l_{it}u_{it}) = 0,\ t = 2, 3, ..., T,\ l = 1, 2, ..., m. \tag{22}$$

Table B summarizes the moment conditions for the level equation.

Table B. Moment Conditions for the Level Equation

Variable	Instruments	Moment conditions
$y_{i,t-1}$	$\Delta y_{i,t-1}$	$E(\Delta y_{i,t-1}u_{it}) = 0,\ t = 2, 3, ..., T$
x^l_{it}	Δx^l_{it}	$E(\Delta x^l_{it}u_{it}) = 0,\ t = 2, 3, ..., T;\ l = 1, 2, ..., m$
w^l_{it}	$\Delta w^l_{i,t-1}$	$E(\Delta w^l_{i,t-1}u_{it}) = 0,\ t = 3, 4, ..., T;\ l = 1, 2, ..., q$

The equation in levels provides $(T-1)$ moment conditions for the lagged dependent variable, $m(T-1)$ 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

$$Y_i = \begin{pmatrix} y_{i0} & 0 & 0 & 0 & 0 & 0 & \cdots 0 \\ 0 & y_{i0} & y_{i1} & 0 & 0 & 0 & \cdots 0 \\ 0 & 0 & 0 & y_{i0} & y_{i1} & y_{i2} & \cdots 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots y_{i0} \cdots y_{i,T_i-2} \end{pmatrix}. \tag{23}$$

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

$$W_i = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & \cdots 0 & 0 \\ w_{i1}^1 & 0 & 0 & \cdots & w_{i1}^q & \cdots 0 & 0 \\ 0 & w_{i1}^1 & w_{i2}^1 & \cdots & 0 & \cdots 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \cdots 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \cdots w_{i,T-3}^q & w_{i,T-2}^q \end{pmatrix}. \tag{24}$$

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.

$$DY_i = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ \Delta y_{i1} & 0 & 0 & \cdots & 0 \\ 0 & \Delta y_{i2} & 0 & \cdots & 0 \\ 0 & 0 & \Delta y_{i3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \Delta y_{i,T-1} \end{pmatrix} \quad DW_i = \begin{pmatrix} 0 & 0\cdots & 0 \\ 0\cdots & 0\cdots & 0 \\ \Delta w_{i2}^1\cdots & \Delta w_{i2}^q\cdots & 0 \\ 0\cdots & 0\cdots & 0 \\ \vdots & \cdots & \vdots \\ 0\cdots & 0\cdots & \Delta w_{i,T-1}^q \end{pmatrix}. \tag{25}$$

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

$$DX_i = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ \Delta x_{i2}^1 & \Delta x_{i2}^2 & \Delta x_{i2}^3 & \cdots & \Delta x_{i2}^m \\ \Delta x_{i3}^1 & \Delta x_{i3}^2 & \Delta x_{i3}^3 & \cdots & \Delta x_{i3}^m \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \Delta x_{iT}^1 & \Delta x_{iT}^2 & \Delta x_{iT}^3 & \cdots & \Delta x_{iT}^m \end{pmatrix} \quad X_i = \begin{pmatrix} x_{i2}^1 & x_{i2}^2 & x_{i2}^3 & \cdots & x_{i2}^m \\ x_{i3}^1 & x_{i3}^2 & x_{i3}^3 & \cdots & x_{i3}^m \\ x_{i4}^1 & x_{i4}^2 & x_{i4}^3 & \cdots & x_{i4}^m \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ x_{iT}^1 & x_{iT}^2 & x_{iT}^3 & \cdots & x_{iT}^m \end{pmatrix}. \tag{26}$$

As shown by Ahn and Schmidt (1995), $(T-1)$ additional linear moment conditions are available if the v_{it} disturbances are assumed to be homoskedastic through time and $E(\Delta y_{i1} u_{i2}) = 0$. Specifically,

$$E(y_{i,t}u_{i,t} - y_{i,t-1}u_{i,t-1}) = 0,\ t = 2, 3, ..., T;\ i = 1, ..., N.$$

Then let Y_i^- be the $T \times T$ instrument matrix used for the moment conditions derived from the homoskedasticity restriction:

$$Y_i^- = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \cdots 0 \\ -y_{i1} & y_{i2} & 0 & 0 & 0 & 0 & \cdots 0 \\ 0 & -y_{i2} & y_{i3} & 0 & 0 & 0 & \cdots 0 \\ 0 & 0 & -y_{i3} & y_{i4} & 0 & 0 & \cdots 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots - y_{i,T-1} \cdots y_{i,T} \end{pmatrix}. \tag{27}$$

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

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

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 (13).

$$u_i = \begin{pmatrix} u_{i1} & u_{i2} & \cdots & u_{iT} \end{pmatrix}' \qquad Dv_i = \begin{pmatrix} \Delta v_{i2} & \Delta v_{i3} & \cdots & \Delta v_{iT} \end{pmatrix}'. \tag{28}$$

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

$$E\left[G_i' U_i\right] = 0 \tag{29}$$

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

$$G_i = \begin{pmatrix} DX_i & DY_i & 0_{T\times T(T-1)/2} & DW_i & 0_{T\times q(T-1)(T-2)/2} & Y_i^- \\ X_i & 0_{(T-1)\times(T-1)} & Y_i & 0_{(T-1)\times q(T-2)} & W_i & 0_{(T-1)\times T} \end{pmatrix}. \tag{30}$$

Based on the moment conditions (29) 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.

3.3 The Limited Information Criterion

As pointed out in section 2.2, evaluating the Bayes factors needed for hypothesis testing and Bayesian model selection or model averaging requires calculating the marginal likelihood

$$p(D|M_j) = \int p(D|\theta, M_j)\, p(\theta|M_j)\, d\theta. \tag{31}$$

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) derives the information projection of the true data generating distribution onto a family of distributions that satisfy certain moment constraints asymptotically. He then uses this information projection directly to build up the quasilikelihood function and he justifies it by showing its large sample properties. Hong and Preston (2008) build a quasi likelihood which is based on objective functions used for extremum estimation (see also Chernozhukov and Hong (2003)). They consider both nested and non-nested model selection and conclude that different penalty functions of model size should be used in each case in order to achieve consistency. However, their quasi-likelihood may change depending on the form of objective functions since their method is only motivated by consistency arguments. In contrast to these approaches, 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). While our approach is related to Schennach (2005) and Ragusa (2008) in spirit, we are able to obtain the likelihood using a much simpler Bayesian procedure by taking advantage of the linear structure of the model, as follows.

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 $\widehat{g}_N(\theta) = N^{-1}\sum_{i=1}^{N} g(\xi_i, \theta)$ denote the sample mean of the moment conditions, and assume that $E\left[g(\xi_i, \theta_0)\, g'(\xi_i, \theta_0)\right]$ and $S(\theta_0) \equiv \lim_{N\rightarrow\infty} Var\left[N^{1/2}\widehat{g}_N(\theta_0)\right]$ 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,* $N^{1/2}\widehat{g}_N(\theta_0) \xrightarrow{d} N(0, S(\theta_0))$.

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

For model (13), the moment conditions for individual i discussed in the previous section can be written in the following form

$$g(\xi_i, \theta) = G_i^{'}(\widetilde{y}_i - \widetilde{z}_i\theta) \tag{32}$$

where $\xi_i = \{\widetilde{y}_i, \widetilde{z}_i\}$, $\widetilde{z}_i = \begin{pmatrix} \widetilde{y}_{i,-1} & \widetilde{x}_i & \widetilde{w}_i \end{pmatrix}$, $\theta = \begin{pmatrix} \alpha & \theta_x & \theta_w \end{pmatrix}^{'}$, while G_i is the matrix defined in (30). The vectors $\widetilde{y}_i$ and $\widetilde{y}_{i,-1}$ for the dependent variable and the lagged dependent variable, respectively, are defined as follows

$$\begin{aligned} \widetilde{y}_i &= \begin{pmatrix} y_{i1} & y_{i2} & \cdots & y_{iT} & \Delta y_{i2} & \Delta y_{i3} & \cdots & \Delta y_{iT} \end{pmatrix}^{'} \\ \widetilde{y}_{i,-1} &= \begin{pmatrix} y_{i0} & y_{i1} & \cdots & y_{i,T-1} & \Delta y_{i1} & \Delta y_{i2} & \cdots & \Delta y_{i,T-1} \end{pmatrix}^{'}. \end{aligned} \tag{33}$$

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

$$\widetilde{x}_i = \begin{pmatrix} x_{i1}^1 & x_{i1}^2 & x_{i1}^3 & \cdots & x_{i1}^m \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ x_{iT}^1 & x_{iT}^2 & x_{iT}^3 & \cdots & x_{iT}^m \\ \Delta x_{i2}^1 & \Delta x_{i2}^2 & \Delta x_{i2}^3 & \cdots & \Delta x_{i2}^m \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \Delta x_{iT}^1 & \Delta x_{iT}^2 & \Delta x_{iT}^3 & \cdots & \Delta x_{iT}^m \end{pmatrix} \tag{34}$$

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

$$\widetilde{w}_i = \begin{pmatrix} w_{i1}^1 & w_{i1}^2 & w_{i1}^3 & \cdots & w_{i1}^q \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ w_{iT}^1 & w_{iT}^2 & w_{iT}^3 & \cdots & w_{iT}^q \\ \Delta w_{i2}^1 & \Delta w_{i2}^2 & \Delta w_{i2}^3 & \cdots & \Delta w_{i2}^q \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \Delta w_{iT}^1 & \Delta w_{iT}^2 & \Delta w_{iT}^3 & \cdots & \Delta w_{iT}^q \end{pmatrix}. \tag{35}$$

Therefore $\widehat{g}_N(\theta_0) = N^{-1}\sum_{i=1}^N G_i^{'}\widetilde{y}_i - N^{-1}\sum_{i=1}^N G'\widetilde{z}_i\theta_0$. By Lemma 1, one may write the likelihood for θ as

$$p\left(N^{-1}\sum_{i=1}^N G_i^{'}\widetilde{y}_i \,\middle|\, \theta, N^{-1}\sum_{i=1}^N G_i^{'}\widetilde{z}_i\right) \propto \exp\left(-\frac{1}{2}N\widehat{g}_N^{'}(\theta)\, S^{-1}(\theta)\widehat{g}_N(\theta)\right). \tag{36}$$

Hence, the model likelihood can be expressed as

$$\int_{\Theta} p\left(N^{-1}\sum_{i=1}^{N} G_i^{'}\widetilde{y}_i|\theta\right) p\left(\theta\right) d\theta \propto \int_{\Theta} \exp\left(-\frac{1}{2}N\widehat{g}_N^{\prime}\left(\theta\right) S^{-1}(\theta)\widehat{g}_N\left(\theta\right)\right) p\left(\theta\right) d\theta.$$

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

$$\int_{\Theta} p\left(N^{-1}\sum_{i=1}^{N} G_i^{'}\widetilde{y}_i|\theta\right) p\left(\theta\right) d\theta \propto \exp\left(\begin{array}{c} -\frac{1}{2}N\widehat{g}_N^{\prime}\left(\widehat{\theta}_0\right) S^{-1}\left(\widehat{\theta}_0\right)\widehat{g}_N\left(\widehat{\theta}_0\right) + \log p\left(\widehat{\theta}_0\right) + \frac{k}{2}\log 2\pi - \\ \frac{1}{2}\log\det\frac{\partial^2}{\partial\theta^2}\left(\frac{1}{2}N\widehat{g}_N^{\prime}\left(\widehat{\theta}_0\right) S^{-1}\left(\hat{\theta}_0\right)\widehat{g}_N\left(\widehat{\theta}_0\right)\right) \end{array}\right)$$

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

$$\int_{\Theta} p\left(N^{-1}\sum_{i=1}^{N} G_i^{'}\widetilde{y}_i|\theta\right) p\left(\theta\right) d\theta \propto \exp\left(-\frac{1}{2}N\widehat{g}_N^{\prime}\left(\widehat{\theta}_0\right) S^{-1}\left(\widehat{\theta}_0\right)\widehat{g}_N\left(\widehat{\theta}_0\right) - \frac{k}{2}\log N\right) \quad (37)$$

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

For a given model M_j for which θ has k_j elements different from zero, with the estimate denoted by $\widehat{\theta}_{0,j}$, the model likelihood (37) becomes

$$\int_{\Theta} p\left(N^{-1}\sum_{i=1}^{N} G_i^{'}\widetilde{y}_i\,|\theta, M_j\right) p\left(\theta\right) d\theta \propto \exp\left(-\frac{1}{2}N\widehat{g}_N^{\prime}\left(\widehat{\theta}_{0,j}\right) S^{-1}\left(\widehat{\theta}_{0,j}\right)\widehat{g}_N\left(\widehat{\theta}_{0,j}\right) - \frac{k_j}{2}\log N\right). \quad (38)$$

Then the moment conditions (29) associated with model M_j can be written as $E\left[G_i^{'}\left(\widetilde{y}_i - \widetilde{z}_i C_{M_j}\theta_0\right)\right] = 0$ where C_{M_j} 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 $\widehat{\theta}_0$ differs from model to model, the sample mean of the moment conditions for model M_j can be written as $\widehat{g}_N\left(\widehat{\theta}_0\right) = N^{-1}\sum_{i=1}^{N} G_i^{'}\left(\widetilde{y}_i - \widetilde{z}_i C_{M_j}\widehat{\theta}_{0,j}\right)$. It is easy to see that $G_i^{'}$, $\widetilde{y}_i$, and $\widetilde{z}_i$ are the same across all models. In other words, the moment conditions and the observable data are the same across the universe of models,[8] allowing us to make valid comparisons of posterior probabilities, in accordance to the principle of Bayesian factor analysis. Therefore, by using (38), one can compute the posterior odds ratio of two models M_1 and M_2 by

---

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

$$\frac{p\left(M_1|N^{-1}\sum_{i=1}^{N}G_i'\widetilde{y}_i\right)}{p\left(M_2|N^{-1}\sum_{i=1}^{N}G_i'\widetilde{y}_i\right)}=\frac{p\left(M_1\right)}{p\left(M_2\right)}\frac{p\left(N^{-1}\sum_{i=1}^{N}G_i'\widetilde{y}_i|M_1\right)}{p\left(N^{-1}\sum_{i=1}^{N}G_i'\widetilde{y}_i|M_2\right)}$$

$$=\frac{p\left(M_1\right)}{p\left(M_2\right)}\exp\left(\begin{array}{c}-\frac{1}{2}\left(N\widehat{g}_N'\left(\widehat{\theta}_{0,1}\right)S^{-1}\left(\widehat{\theta}_{0,1}\right)\widehat{g}_N\left(\widehat{\theta}_{0,1}\right)-N\widehat{g}_N'\left(\widehat{\theta}_{0,2}\right)S^{-1}\left(\widehat{\theta}_{0,2}\right)\widehat{g}_N\left(\widehat{\theta}_{0,2}\right)\right)\\ -\left(\frac{k_1-k_2}{2}\log N\right)\end{array}\right) \tag{39}$$

which has the same form of BIC as fully specified models. We use iterative GMM estimation with moment conditions $E\left[G_i'\left(\widetilde{y}_i-\widetilde{z}_iC_{M_j}\theta_{0,j}\right)\right]=0$ to approximate the Bayesian factors above. A consistent estimate of the weighting matrix is used to replace $S^{-1}\left(\widehat{\theta}_0\right)$ in (39). In our simulations, we assume a unit information prior for the parameters (as discussed in Section 2) and a Uniform distribution over the model space, essentially implying that there is no preference for a specific model so $p(M_1)=p(M_2)=...=p(M_K)=\frac{1}{K}$.

4 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.

4.1 The Data Generating Process

We consider the case where the universe of potential explanatory variables contains 9 variables, namely, 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.

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

$$\begin{array}{c}\left(\begin{array}{cccc}x_{it}^1 & x_{it}^2 & x_{it}^3 & x_{it}^4\end{array}\right)=\left(\begin{array}{cccc}0.3 & 0.4 & 0.8 & 0.5\end{array}\right)+r_t\\ \text{with } r_t\sim N\left(0,I_4\right)\text{ for } t=0,1,...,T;\ i=1,...,N,\end{array} \tag{40}$$

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, $\left(\begin{array}{cc}x_i^5 & x_i^6\end{array}\right)$ are correlated with $\left(\begin{array}{cc}x_i^1 & x_i^2\end{array}\right)$ such that for every individual i and period t, the data generating process is given by

$$\begin{array}{l}\left(\begin{array}{cc} x_{it}^5 & x_{it}^6 \end{array}\right) = \left(\left(\begin{array}{cc} x_{it}^1 & x_{it}^2 \end{array}\right) - \left(\begin{array}{cc} 0.3 & 0.4 \end{array}\right)\right) \cdot 0.1 \cdot \left(\begin{array}{cc} 1 & 2 \end{array}\right)' \left(\begin{array}{cc} 1 & 1 \end{array}\right) + \left(\begin{array}{cc} 1.5 & 1.8 \end{array}\right) + r_t \\ \text{with } r_t \sim N\left(0, I_2\right) \text{ for } t = 0, 1, ..., T;\ i = 1, ..., N, \end{array} \quad (41)$$

where I_2 is the two dimensional identity matrix.

Similarly, for the endogenous variables, $\left(\begin{array}{cc} w_i^1 & w_i^2 \end{array}\right)$, we have the following data generating process

$$\begin{array}{l} \left(\begin{array}{cc} w_{it}^1 & w_{it}^2 \end{array}\right) = 0.71 \left(\begin{array}{cc} w_{i,t-1}^1 & w_{i,t-1}^2 \end{array}\right) + 6.7 v_{it} \left(\begin{array}{cc} 1 & 1 \end{array}\right) + r_t \text{ for } t = 1, 2, ..., T \\ \left(\begin{array}{cc} w_{i0}^1 & w_{i0}^2 \end{array}\right) = 6.7 v_{i0} \left(\begin{array}{cc} 1 & 1 \end{array}\right) + r_0 \\ \text{with } v_{it} \sim N\left(0, \sigma_v^2\right) \text{ and } r_t \sim N\left(0, I_2\right) \text{ for } t = 0, 1, ..., T. \end{array} \quad (42)$$

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

$$\begin{array}{l} y_{i0} = \frac{1}{(1-\alpha)} \left(x_{i0}\theta_x + w_{i0}\theta_w + \eta_i + v_{i0}\right) \\ \text{with } v_{i0} \sim N\left(0, \sigma_v^2\right) \text{ and } \eta_i \sim N\left(0, \sigma_\eta^2\right) \end{array} \quad (43)$$

where $\theta_x = \left(\begin{array}{cccccc} 0.05 & 0 & 0 & -0.05 & 0 & 0.05 \end{array}\right)'$, $\theta_w = \left(\begin{array}{cc} 0 & 0.13 \end{array}\right)'$, $w_{i0} = \left(\begin{array}{cc} w_{i0}^1 & w_{i0}^2 \end{array}\right)$, and $x_{i0} = \left(\begin{array}{cccccc} x_{i0}^1 & x_{i0}^2 & x_{i0}^3 & x_{i0}^4 & x_{i0}^5 & x_{i0}^6 \end{array}\right)$.

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

$$\begin{array}{l} y_{it} = \alpha y_{i,t-1} + \theta_x x_{it} + \theta_w w_{it} + \eta_i + v_{it} \\ \text{with } v_{it} \sim N\left(0, \sigma_v^2\right) \text{ and } \eta_i \sim N\left(0, \sigma_\eta^2\right). \end{array} \quad (44)$$

The theoretical R^2 of the generated model varies between 0.50 and 0.60.

We test the robustness of our procedure with respect to the underlying distributions of the error term by relaxing the normality assumption and using discrete distributions instead. Concretely, the distribution of the random variable v_{it}, is obtained in the following way. 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 $\omega_k \sim \mathit{Exponential}\,(1)$. 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 adjust each point in S_v so that v_{it} has zero mean and variance σ_v^2. 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}.

4.2 Simulation Results

This section reports Monte Carlo simulations of our LIBMA methodology in order to assess its performance. We generate 1000 instances of the data generating process with the exogenous variables x_{it}, endogenous variables w_{it}, and parameter values $\begin{pmatrix} \alpha & \theta_x & \theta_w \end{pmatrix}'$ as discussed in the previous section, and we present results in the form of medians, means, variances and quartiles. We consider several sample sizes, $N = 200,\ 500,$ and 2000, and two values for the coefficient of the lagged dependent variable, $\alpha = 0.95$ and 0.50. 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, $v_{it} \sim N\left(0, \sigma_v^2\right)$ and $\eta_i \sim N\left(0, \sigma_\eta^2\right)$, respectively. We consider the cases where $\sigma_v^2 = 0.05,\ 0.10,$ and 0.20 while $\sigma_\eta^2 = 0.10$. 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.

4.2.1 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 1000 instances. As expected, the mean posterior probabilities of the true model increase with the sample size. For sample size $N = 200$ the mean of the posterior probability of the true model ranges from 0.031 to 0.218, depending on the values of the other parameters. As the sample size increases to $N = 2000$, the mean of the posterior probability of the true model increases, while at the same time, showing less variation across different combinations of parameters, with the values ranging from 0.633 to 0.655. As the sample size increases the median posterior model probabilities become slightly higher than the means, ranging from 0.690 to 0.705 for $N = 2000$. In addition, as the sample increases, the distribution of the posterior probabilities of the true model becomes skewed toward 1, as shown by the quartiles in Table 1 and the density plots in Figure 1, Appendix A.[9]

It is easy to see from equation (3) that 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 compute a relative measure that helps one understand how well the methodology performs, independent of the size of the universe. 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). This ratio would be above 1 if the true model has the highest posterior probability and below 1 if there exists another model with a higher posterior probability than the true model. For sample sizes $N = 500$ and above, this ratio is above unity for all the cases considered, suggesting that the correct model is on average favored over all the other models. For the smaller sample, $N = 200$, the ratio decreases from to 1.591 and 1.039 to 0.422 and 0.249, respectively, as the variance of the random error term increases from 0.05 to 0.20. As expected, the average ratios increase

---

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

with the sample size, reaching values above 6.5 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. For the smallest sample size, $N = 200$, the recovery rate varies from 7 percent to 59 percent and it decreases as the variance of the random error term increases from 0.05 to 0.20. For $N = 500$ we see an improvement in the selection of the true model with the success rate ranging from 51 percent to 83 percent. The variation becomes much smaller for $N = 2000$ with the recovery rate ranging from 91 to 94 percent.

4.2.2 Model averaging

While model selection properties are desired, the strength of our methodology is given by its performance in the context of Bayesian Model Averaging. Sometimes researchers may not be interested in recovering the exact data generating process, but rather understand which of the variables under consideration are more likely to belong to the true model. That is the area where the BMA may help us. One measure that we report for our experiments is the inclusion probability for each variable considered. The inclusion probability for a given variable is defined as the sum of all the posterior probabilities for each model that contains that particular variable. Table 4 presents the posterior inclusion probabilities for all the variables considered along with the true model (column 2 of the table).[10] 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, for samples $N \geq 500$, the median value of the inclusion probability for all the relevant explanatory variables is greater than 0.95 in all cases considered. As the sample size increases the posterior inclusion probabilities approach 1 for all the relevant variables. 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 0.07 for all the cases considered when $N = 2000$. It is interesting to see that even in cases where the recovery rate of the true model is poor (12 percent for the case in which $N = 200, \alpha = 0.95, \sigma_v^2 = 0.20$), the probability of inclusion is able to differentiate among the relevant and non-relevant variables.

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 1000 replications, compared to the parameters of the true model.[11] As in the case of inclusion probabilities, our methodology is performing 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 1000 replications is also very small across the board with values less than 10^{-4} in many cases.

---

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

[11] 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.

4.2.3 Robustness checks using non-Gaussian errors

As discussed above 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, we see again that the mean posterior probabilities of the true model increase with the sample size, while at the same time, showing less variation across different combinations of parameters. Moreover, the sample size increases, the median posterior model probabilities become slightly higher than the means, ranging from 0.684 to 0.708 for $N = 2000$. In addition, the distribution of the posterior probabilities of the true model becomes skewed toward 1, as shown by the quartiles in Table 6 and the density plots in Figure 4. For Table 7, conclusions are similar to Table 2. For sample sizes $N = 500$ and above, the ratio of the posterior model probability of the true model to the highest posterior probability of all the other models is above unity for all the cases considered, suggesting that the correct model is on average favored over all the other models. For the smaller sample, $N = 200$, the ratio decreases from to 1.587 and 1.205 to 0.350 and 0.254, respectively, as the variance of the random error term increases from 0.05 to 0.20. As expected, the average ratios increase with the sample size, reaching values above 6.4 for $N = 2000$.

Model recovery under non-Gaussian errors is still 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 7 percent to 59 percent and it decreases as the variance of the random error term increases from 0.05 to 0.20. For $N = 500$ we see an improvement in the selection of the true model with the success rate ranging from 51 percent to 85 percent. The variation becomes much smaller for $N = 2000$ with the recovery rate ranging from 92 to 93 percent.

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, for samples $N \geq 500$, the median value of the inclusion probability for all the relevant explanatory variables is greater than 0.90 in all cases considered. As the sample size increases the posterior inclusion probabilities approach 1 for all the relevant variables. 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 0.073 for all the cases considered when $N = 2000$. It is interesting to see that even in cases where the recovery rate of the true model is poor (8 percent for the case in which $N = 200, \alpha = 0.95, \sigma_v^2 = 0.20$), the probability of inclusion is able to differentiate among the relevant and non-relevant variables. In Table 10, estimated parameter medians and variances are very close to those reported in Table 5. As in the Gaussian case, our methodology is performing well in estimating the parameters, with the performance improving as the sample gets larger. In Figure 6 of Appendix A we present the box plots for the parameter estimates of Table 10 (for the case of $\alpha = 0.95$ and $\sigma_v = 0.1$). Again, the variance of the distribution decreases as the sample size increases and the median moves toward the true value.

5 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 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 dynamic panel data models with endogenous regressors.

Future research could 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

[1] Ahn, S.C., Schmidt, P., 1995. Efficient estimation of models for dynamic panel data. *Journal of Econometrics,* Vol· 68, pp. 5-28.

[2] 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.

[3] Arellano, M., Bond, S.R., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58, pp. 277-297.

[4] Arellano, M., Bover, O., 1995. Another look at the instrumental-variable estimation of error components models. *Journal of Econometrics,* Vol. 68, pp. 29-52.

[5] Back, K. and D. P. Brown, 1993, "Implied Probabilities in GMM Estimators", *Econometrica*, Vol. 61, No. 4, pp. 971–975.

[6] Blundell, R., and S. Bond, 1998, "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, Vol. 87, No. 1, pp. 114–143.

[7] Brock, W., and S. Durlauf, 2001, "Growth Empirics and Reality," *World Bank Economic Review*, 15, pp. 229–272.

[8] Chernozhukov, V., and H. Hong, 2003, "An MCMC Approach to Classical Estimation," *Journal of Econometrics*, Vol. 115, No. 2, pp. 293–346.

[9] 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.

[10] 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).

[11] Fernàndez C., E. Ley, and M. Steel, 2001a, "Model Uncertainty in Cross-Country Growth Regressions," *Journal of Applied Econometrics*, Vol. 16, pp. 563–76.

[12] Fernàndez C., E. Ley and M.F.J. Steel, 2001b, "Benchmark Priors for Bayesian Model Averaging," *Journal of Econometrics*, Vol. 100, pp. 381–427.

[13] Hall, A.R., 2005, Generalized Method of Moments (New York: Oxford University Press).

[14] 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.

[15] Hong, H., and B. Preston, 2008, "Bayesian Averaging, Prediction and Nonnested Model Selection," *NBER Working Papers 14284*, National Bureau of Economic Research, Inc.

[16] Jacobson T., and S. Karlsson, 2004, "Finding Good Predictors for Inflation: A Bayesian Model Averaging Approach", *Journal of Forecasting*, Vol. 23, pp. 476–496.

[17] Kass, R., and A. Raftery, 1995, "Bayes Factors," *Journal of the American Statistical Association*, Vol. 90, No. 430, pp. 773–95.

[18] 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.

[19] Kim, J. Y., 2002, "Limited Information Likelihood and Bayesian Analysis," *Journal of Econometrics*, Vol. 107, No. 1-2, pp. 175–93.

[20] 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.

[21] Leamer, E., 1978, Specification Searches: Ad Hoc Inference with Non-experimental Data (New York: Wiley).

[22] Leamer, E., 1983, "Let's Take the Con Out of Econometrics," *American Economic Review*, Vol. 73, pp. 31–43.

[23] 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.

[24] 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.

[25] 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.

[26] Moral-Benito, E., 2007, "Determinants of Economic Growth: A Bayesian Panel Data Approach," CEMFI Working Papers Paper 0719.

[27] 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.

[28] 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.

[29] Raftery, A. E., 1995, "Bayesian Model Selection in Social Research," *Sociological Methodology*, Vol. 25, pp. 111–163.

[30] Raftery, A. E., 1996, "Approximate Bayes Factors and Accounting for Model Uncertainty in Generalized Linear Models," *Biometrika*, Vol. 83, pp. 251–66.

[31] 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.

[32] Ragusa, G., 2008, "Bayesian Likelihoods for Moment Condition Models," mimeo.

[33] Sala-i-Martin, X., Doppelhofer, G., R., and placeI. 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.

[34] Schennach, StateS.C., 2005, "Bayesian Exponentially Tilted Empirical Likelihood," *Biometrika*, Vol. 92, pp. 31–46.

[35] Tsangarides, C., 2004, "A Bayesian Approach to Model Uncertainty," IMF Working Paper No. 04/68 (Washington: International Monetary Fund).

[36] 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.

[37] 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.

Tables

Table 1. Posterior probability of the true model

Summary statistics using LIBMA estimation for various N, α, and σ_v

α	0.95			0.50		
σ_v	0.05	0.10	0.20	0.05	0.10	0.20
Sample						
N=200						
Mean	0.218	0.112	0.052	0.140	0.094	0.031
Variance	0.021	0.011	0.005	0.023	0.016	0.004
Q1	0.091	0.028	0.010	0.008	0.003	0.000
Median	0.213	0.078	0.027	0.077	0.033	0.004
Q3	0.340	0.169	0.066	0.257	0.141	0.030
N=500						
Mean	0.448	0.419	0.275	0.429	0.403	0.257
Variance	0.025	0.025	0.027	0.030	0.033	0.035
Q1	0.363	0.310	0.132	0.319	0.264	0.085
Median	0.485	0.455	0.266	0.466	0.440	0.234
Q3	0.574	0.547	0.412	0.568	0.562	0.420
N=2000						
Mean	0.633	0.652	0.655	0.646	0.646	0.644
Variance	0.027	0.023	0.020	0.023	0.027	0.025
Q1	0.585	0.603	0.609	0.584	0.601	0.588
Median	0.690	0.705	0.702	0.699	0.702	0.694
Q3	0.747	0.757	0.753	0.754	0.757	0.753

Notes:

1. For the idiosyncratic error term, $\eta_i \sim N(0,\sigma_\eta^2)$ where $\sigma_\eta^2 = 0.10$.
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

α	0.95			0.50		
σ_v	0.05	0.10	0.20	0.05	0.10	0.20
Sample						
N=200						
Mean	1.591	0.856	0.422	1.039	0.761	0.249
Variance	1.757	0.922	0.438	1.661	1.348	0.312
Q1	0.443	0.189	0.074	0.039	0.017	0.001
Median	1.259	0.503	0.189	0.425	0.220	0.024
Q3	2.480	1.195	0.436	1.726	1.012	0.210
N=500						
Mean	3.254	3.113	1.975	3.034	2.965	1.812
Variance	4.709	4.618	3.286	4.735	5.073	3.647
Q1	1.407	1.313	0.541	1.146	1.024	0.328
Median	2.977	2.788	1.402	2.687	2.526	1.054
Q3	4.910	4.632	2.953	4.670	4.571	2.832
N=2000						
Mean	6.534	7.164	7.030	6.930	6.990	6.770
Variance	17.875	19.696	18.042	19.164	19.708	18.381
Q1	3.066	3.348	3.440	3.096	3.316	3.185
Median	6.040	6.854	6.728	6.538	6.414	6.128
Q3	9.424	10.634	10.298	10.380	10.308	10.054

Notes:

1. For the idiosyncratic error term, $\eta_i \sim N(0,\sigma_\eta^2)$ where $\sigma_\eta^2 = 0.10$.
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

α	0.95			0.50		
σ_v	0.05	0.10	0.20	0.05	0.10	0.20
Sample						
N=200						
% Correct	59	29	12	35	25	7
N=500						
% Correct	83	80	59	78	76	51
N=2000						
% Correct	91	93	94	93	93	93

Notes:

1. For the idiosyncratic error term, $\eta_i \sim N(0,\sigma_\eta^2)$ where $\sigma_\eta^2 = 0.10$.

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

α		0.95						0.50					
σ_v	True	0.05		0.10		0.20		0.05		0.10		0.20	
Sample	Model	Median	Variance	Median	Variance	Median	Variance	Median	Variance	Median	Variance	Median	Variance
N=200													
y_{t-1}	1	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00213	1.00000	0.00001	1.00000	0.00000
x_1	1	0.98188	0.03253	0.82971	0.06918	0.57772	0.08462	0.91774	0.06038	0.75181	0.08323	0.45270	0.08127
x_2	0	0.18687	0.01733	0.18811	0.02209	0.18926	0.02062	0.19268	0.02115	0.18616	0.01964	0.19081	0.02260
x_3	0	0.18515	0.01881	0.18749	0.02091	0.18702	0.01854	0.18788	0.01856	0.18887	0.02124	0.18820	0.02283
x_4	1	0.97537	0.03739	0.81329	0.07214	0.53740	0.08149	0.90190	0.06915	0.65897	0.08478	0.45905	0.08000
x_5	0	0.18911	0.01907	0.18589	0.02087	0.18596	0.01972	0.18754	0.01960	0.18482	0.01892	0.19068	0.01738
x_6	1	0.97151	0.04275	0.76711	0.07494	0.56244	0.08735	0.91728	0.06382	0.64502	0.08496	0.41319	0.07949
w_1	0	0.19034	0.04216	0.21460	0.05148	0.26059	0.05947	0.12571	0.04075	0.10747	0.03675	0.07218	0.04837
w_2	1	0.98348	0.07007	0.97228	0.07305	0.94732	0.07147	0.53306	0.15868	0.35818	0.14767	0.09843	0.12655
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
x_1	1	1.00000	0.00001	0.99976	0.00516	0.96843	0.04591	1.00000	0.00006	0.99987	0.00852	0.98045	0.05822
x_2	0	0.12874	0.01732	0.12770	0.01525	0.12823	0.01838	0.13179	0.02375	0.13112	0.02293	0.13124	0.02793
x_3	0	0.12776	0.01952	0.12561	0.01471	0.12810	0.01802	0.12923	0.02274	0.12662	0.02196	0.12764	0.02379
x_4	1	1.00000	0.00013	0.99971	0.00510	0.96040	0.04735	1.00000	0.00060	0.99983	0.00766	0.97808	0.05766
x_5	0	0.12404	0.01551	0.12323	0.01693	0.12605	0.02026	0.12934	0.02120	0.12611	0.02471	0.12596	0.02376
x_6	1	1.00000	0.00054	0.99920	0.00990	0.95687	0.04778	1.00000	0.00037	0.99958	0.01071	0.95938	0.06335
w_1	0	0.12121	0.02613	0.12725	0.02709	0.12507	0.02530	0.10749	0.02368	0.10812	0.02581	0.10405	0.03400
w_2	1	1.00000	0.00003	1.00000	0.00054	0.99995	0.00640	0.99994	0.01551	0.99976	0.02369	0.98946	0.09191
N=2000													
y_{t-1}	1	1.00000	0.00000	1.00000	0.00100	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	0.00100	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000
x_2	0	0.06983	0.01262	0.06461	0.01118	0.06679	0.00756	0.06561	0.01151	0.06467	0.01161	0.06677	0.01262
x_3	0	0.06471	0.00800	0.06498	0.00915	0.06619	0.00924	0.06797	0.00949	0.06743	0.01418	0.06703	0.01095
x_4	1	1.00000	0.00000	1.00000	0.00100	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000
x_5	0	0.06782	0.01047	0.06488	0.00900	0.06533	0.01144	0.06708	0.01234	0.06496	0.01137	0.06666	0.01200
x_6	1	1.00000	0.00000	1.00000	0.00100	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00002
w_1	0	0.06897	0.02023	0.06684	0.01238	0.06546	0.00899	0.06235	0.01077	0.06223	0.01272	0.06308	0.01090
w_2	1	1.00000	0.00000	1.00000	0.00100	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000

Notes:

1. For the idiosyncratic error term, $\eta_i \sim N(0,\sigma_\eta^2)$ where $\sigma_\eta^2 = 0.10$.
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						
σ_v	True	0.05		0.10		0.20		True	0.05		0.10		0.20	
Sample	Value	Median	Variance	Median	Variance	Median	Variance	Value	Median	Variance	Median	Variance	Median	Variance
N=200														
y_{t-1}	0.95	0.95174	0.00003	0.95077	0.00001	0.94994	0.00001	0.50	0.57711	0.02219	0.58787	0.01905	0.64631	0.01903
x_1	0.05	0.04854	0.00027	0.04246	0.00056	0.03164	0.00085	0.05	0.04401	0.00047	0.03949	0.00078	0.02613	0.00114
x_2	0.00	0.00048	0.00004	0.00047	0.00009	0.00038	0.00016	0.00	-0.00002	0.00007	-0.00016	0.00012	0.00102	0.00031
x_3	0.00	0.00007	0.00004	-0.00017	0.00008	-0.00013	0.00014	0.00	0.00031	0.00007	0.00025	0.00012	-0.00027	0.00036
x_4	-0.05	-0.04656	0.00029	-0.04049	0.00056	-0.02747	0.00078	-0.05	-0.04312	0.00049	-0.03461	0.00066	-0.02707	0.00108
x_5	0.00	0.00016	0.00004	-0.00010	0.00009	0.00022	0.00015	0.00	0.00006	0.00006	-0.00007	0.00015	0.00041	0.00026
x_6	0.05	0.04800	0.00034	0.03840	0.00061	0.03123	0.00096	0.05	0.04673	0.00056	0.03463	0.00075	0.02460	0.00117
w_1	0.00	0.00264	0.00058	0.00349	0.00105	0.00537	0.00171	0.00	0.00246	0.00047	0.00231	0.00059	0.00200	0.00103
w_2	0.13	0.14005	0.00194	0.13819	0.00205	0.13350	0.00217	0.13	0.07728	0.00477	0.06154	0.00482	0.01653	0.00424
N=500														
y_{t-1}	0.95	0.95115	0.00001	0.95064	0.00000	0.94995	0.00000	0.50	0.47413	0.00244	0.45469	0.00307	0.46474	0.00807
x_1	0.05	0.04975	0.00006	0.05085	0.00014	0.04968	0.00039	0.05	0.04920	0.00006	0.04981	0.00013	0.04807	0.00042
x_2	0.00	0.00015	0.00001	0.00015	0.00002	0.00012	0.00004	0.00	-0.00012	0.00001	-0.00006	0.00002	0.00014	0.00007
x_3	0.00	-0.00001	0.00001	-0.00004	0.00002	-0.00005	0.00004	0.00	0.00011	0.00001	0.00009	0.00002	-0.00010	0.00005
x_4	-0.05	-0.04895	0.00006	-0.05049	0.00012	-0.04799	0.00037	-0.05	-0.04858	0.00006	-0.04915	0.00012	-0.04727	0.00040
x_5	0.00	0.00012	0.00001	0.00001	0.00002	0.00015	0.00005	0.00	-0.00002	0.00001	0.00003	0.00002	0.00010	0.00005
x_6	0.05	0.05021	0.00007	0.05045	0.00017	0.04958	0.00041	0.05	0.04977	0.00006	0.04900	0.00015	0.04701	0.00044
w_1	0.00	0.00073	0.00006	0.00065	0.00010	0.00025	0.00024	0.00	0.00087	0.00005	0.00078	0.00011	0.00122	0.00039
w_2	0.13	0.13953	0.00012	0.13727	0.00015	0.13647	0.00031	0.13	0.14568	0.00057	0.15412	0.00082	0.14617	0.00251
N=2000														
y_{t-1}	0.95	0.95018	0.00000	0.95017	0.00090	0.94997	0.00000	0.50	0.49532	0.00025	0.47973	0.00022	0.48687	0.00027
x_1	0.05	0.05050	0.00001	0.05140	0.00003	0.05169	0.00005	0.05	0.04966	0.00001	0.05096	0.00002	0.05046	0.00005
x_2	0.00	0.00008	0.00000	0.00005	0.00000	0.00000	0.00000	0.00	-0.00006	0.00000	-0.00004	0.00000	0.00000	0.00001
x_3	0.00	0.00000	0.00000	-0.00004	0.00000	-0.00002	0.00000	0.00	0.00007	0.00000	0.00010	0.00000	-0.00004	0.00001
x_4	-0.05	-0.04874	0.00001	-0.05020	0.00003	-0.04985	0.00005	-0.05	-0.04862	0.00001	-0.05010	0.00002	-0.05174	0.00005
x_5	0.00	0.00003	0.00000	-0.00001	0.00000	0.00007	0.00001	0.00	-0.00002	0.00000	0.00003	0.00000	0.00003	0.00001
x_6	0.05	0.05010	0.00002	0.05070	0.00003	0.05126	0.00006	0.05	0.05029	0.00002	0.04905	0.00003	0.04933	0.00005
w_1	0.00	0.00030	0.00001	0.00024	0.00001	0.00000	0.00001	0.00	0.00020	0.00001	0.00009	0.00001	0.00033	0.00001
w_2	0.13	0.13280	0.00003	0.13114	0.00004	0.13202	0.00002	0.13	0.13285	0.00006	0.14081	0.00006	0.13539	0.00007

Notes:

1. For the idiosyncratic error term, $\eta_i \sim N(0,\sigma_\eta^2)$ where $\sigma_\eta^2 = 0.10$.
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

α	0.95			0.50		
σ_v	0.05	0.10	0.20	0.05	0.10	0.20
Sample						
N=200						
Mean	0.224	0.109	0.044	0.162	0.084	0.031
Variance	0.021	0.012	0.004	0.025	0.015	0.005
Q1	0.094	0.025	0.008	0.019	0.002	0.000
Median	0.222	0.071	0.023	0.115	0.023	0.003
Q3	0.342	0.161	0.053	0.283	0.118	0.023
N=500						
Mean	0.459	0.415	0.253	0.437	0.376	0.243
Variance	0.023	0.025	0.028	0.032	0.036	0.035
Q1	0.371	0.315	0.110	0.322	0.233	0.069
Median	0.497	0.447	0.231	0.487	0.412	0.211
Q3	0.580	0.543	0.386	0.582	0.534	0.397
N=2000						
Mean	0.653	0.632	0.643	0.640	0.643	0.651
Variance	0.021	0.025	0.025	0.026	0.025	0.024
Q1	0.607	0.573	0.585	0.592	0.595	0.598
Median	0.703	0.684	0.703	0.700	0.699	0.708
Q3	0.754	0.747	0.751	0.753	0.754	0.757

Notes:
1. The error terms are constructed using discrete distributions (see Section 4.1.).

Table 7. Posterior probability ratio: true model/best among the other models
Summary statistics using LIBMA estimation for various N, α, and σ_v

α	0.95			0.50		
σ_v	0.05	0.10	0.20	0.05	0.10	0.20
Sample						
N=200						
Mean	1.587	0.828	0.350	1.205	0.662	0.254
Variance	1.767	1.043	0.348	1.813	1.168	0.420
Q1	0.463	0.168	0.053	0.099	0.011	0.001
Median	1.235	0.423	0.173	0.639	0.141	0.016
Q3	2.499	1.081	0.357	1.975	0.812	0.162
N=500						
Mean	3.384	3.050	1.783	3.258	2.721	1.715
Variance	4.835	4.355	3.257	5.489	4.810	3.590
Q1	1.532	1.318	0.437	1.205	0.867	0.258
Median	3.065	2.705	1.118	2.862	2.216	1.044
Q3	5.035	4.584	2.630	5.069	4.278	2.487
N=2000						
Mean	6.980	6.410	6.992	6.828	6.824	7.081
Variance	18.234	17.732	19.169	19.496	18.475	19.192
Q1	3.360	2.956	3.035	3.138	3.257	3.221
Median	6.676	5.738	6.824	6.335	6.464	6.762
Q3	10.195	9.520	10.229	9.999	10.021	10.578

Notes:
1. The error terms are constructed using discrete distributions (see Section 4.1.).

Table 8. Probability of retrieving the true model
Summary statistics using LIBMA estimation for various n, α, and σ_v

α	0.95			0.50		
σ_v	0.05	0.10	0.20	0.05	0.10	0.20
Sample						
N=200						
% Correct	56	27	8	42	22	7
N=500						
% Correct	85	80	53	79	72	51
N=2000						
% Correct	93	92	93	92	92	93

Notes:

1. The error terms are constructed using discrete distributions (see Section 4.1.).

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

α		0.95						0.50					
σ_v	True	0.05		0.10		0.20		0.05		0.10		0.20	
Sample	Model	Median	Variance	Median	Variance	Median	Variance	Median	Variance	Median	Variance	Median	Variance
N=200													
y_{t-1}	1	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00103	1.00000	0.00000	1.00000	0.00001
x_1	1	0.98533	0.03332	0.84439	0.07089	0.54259	0.08866	0.94050	0.05592	0.69674	0.08352	0.46005	0.08605
x_2	0	0.18906	0.01873	0.19277	0.02222	0.18649	0.02084	0.18470	0.01901	0.19136	0.02213	0.18425	0.02000
x_3	0	0.18712	0.01748	0.18462	0.02122	0.17957	0.01710	0.18935	0.02013	0.18396	0.01668	0.18991	0.02314
x_4	1	0.98708	0.02926	0.83962	0.07696	0.49877	0.08407	0.94699	0.05413	0.63426	0.08962	0.37623	0.07540
x_5	0	0.18621	0.01949	0.18930	0.02272	0.18612	0.02089	0.18920	0.01543	0.19009	0.01948	0.18938	0.02186
x_6	1	0.96697	0.04240	0.78891	0.08121	0.47078	0.08273	0.91206	0.05738	0.64015	0.08609	0.33444	0.07223
w_1	0	0.19013	0.04195	0.21127	0.05684	0.26817	0.05852	0.13175	0.03713	0.10253	0.04647	0.06670	0.04599
w_2	1	0.98807	0.05995	0.96496	0.07656	0.95082	0.06951	0.64884	0.15207	0.27934	0.15240	0.09806	0.12459
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
x_1	1	1.00000	0.00000	0.99975	0.00718	0.96429	0.04964	1.00000	0.00017	0.99984	0.01012	0.98394	0.04787
x_2	0	0.12575	0.02023	0.12433	0.01388	0.12599	0.01826	0.12894	0.02064	0.12785	0.02530	0.12986	0.02306
x_3	0	0.12148	0.01611	0.12167	0.01720	0.12262	0.01699	0.12757	0.02872	0.13145	0.02755	0.12779	0.02544
x_4	1	1.00000	0.00003	0.99964	0.00677	0.95426	0.05240	1.00000	0.00041	0.99959	0.01251	0.96482	0.06551
x_5	0	0.12695	0.01818	0.12891	0.01844	0.12531	0.01848	0.12775	0.02117	0.13564	0.02351	0.12837	0.02320
x_6	1	1.00000	0.00025	0.99865	0.01006	0.93790	0.06465	1.00000	0.00151	0.99912	0.01641	0.90998	0.08230
w_1	0	0.11931	0.01722	0.12270	0.02526	0.12990	0.02686	0.10587	0.02217	0.10632	0.02910	0.10812	0.03094
w_2	1	1.00000	0.00000	1.00000	0.00059	0.99995	0.00698	0.99997	0.01091	0.99912	0.04088	0.98931	0.08415
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
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
x_2	0	0.06637	0.00923	0.06548	0.00837	0.06676	0.01228	0.06620	0.01005	0.06804	0.01255	0.06577	0.01074
x_3	0	0.06416	0.01034	0.06834	0.01250	0.06516	0.00777	0.06966	0.01613	0.06630	0.00990	0.06572	0.00927
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	0.00002
x_5	0	0.06750	0.00914	0.07325	0.01962	0.06616	0.01124	0.06694	0.01088	0.06875	0.01403	0.06504	0.01124
x_6	1	1.00000	0.00000	1.00000	0.00000	1.00000	0.00001	1.00000	0.00000	1.00000	0.00000	1.00000	0.00005
w_1	0	0.06465	0.01128	0.06403	0.00890	0.06718	0.01530	0.06199	0.01127	0.06351	0.01194	0.06393	0.01324
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

Notes:

1. The error terms are constructed using discrete distributions (see Section 4.1.).

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

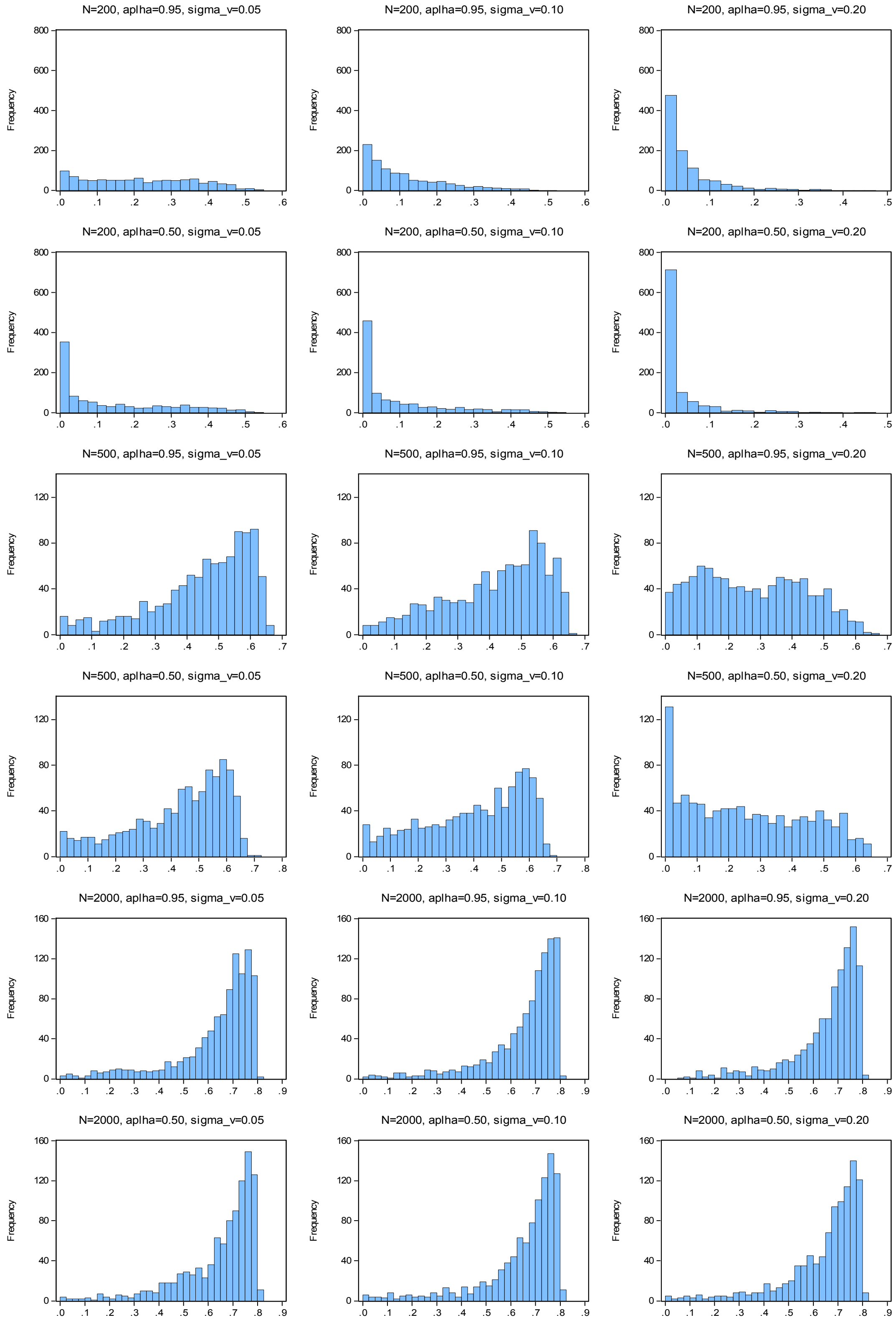
α	0.95							0.50						
σ_v	True	0.05		0.10		0.20		True	0.05		0.10		0.20	
Sample	Value	Median	Variance	Median	Variance	Median	Variance	Value	Median	Variance	Median	Variance	Median	Variance
N=200														
y_{t-1}	0.95	0.95265	0.00003	0.95088	0.00001	0.95004	0.00001	0.50	0.54517	0.02128	0.58207	0.01934	0.64121	0.01742
x_1	0.05	0.04877	0.00029	0.04260	0.00059	0.02708	0.00092	0.05	0.04642	0.00044	0.03736	0.00076	0.02658	0.00135
x_2	0.00	0.00000	0.00004	0.00019	0.00009	-0.00012	0.00016	0.00	0.00034	0.00006	0.00031	0.00017	0.00056	0.00029
x_3	0.00	0.00003	0.00004	0.00024	0.00008	-0.00006	0.00012	0.00	-0.00038	0.00007	-0.00021	0.00011	0.00031	0.00038
x_4	-0.05	-0.05008	0.00028	-0.04235	0.00060	-0.02513	0.00083	-0.05	-0.04809	0.00045	-0.03288	0.00076	-0.02041	0.00094
x_5	0.00	-0.00019	0.00004	0.00065	0.00009	0.00011	0.00017	0.00	0.00004	0.00005	0.00051	0.00014	-0.00025	0.00039
x_6	0.05	0.04781	0.00035	0.04159	0.00066	0.02405	0.00089	0.05	0.04483	0.00048	0.03542	0.00087	0.01822	0.00102
w_1	0.00	0.00241	0.00053	0.00393	0.00124	0.00591	0.00173	0.00	0.00245	0.00034	0.00232	0.00082	0.00164	0.00104
w_2	0.13	0.14234	0.00168	0.13564	0.00206	0.13346	0.00206	0.13	0.10195	0.00457	0.04429	0.00481	0.01747	0.00432
N=500														
y_{t-1}	0.95	0.95153	0.00001	0.95077	0.00000	0.95017	0.00000	0.50	0.47206	0.00211	0.46203	0.00390	0.46398	0.00707
x_1	0.05	0.05043	0.00006	0.05011	0.00014	0.04933	0.00042	0.05	0.04855	0.00005	0.04995	0.00015	0.05036	0.00039
x_2	0.00	-0.00007	0.00001	0.00006	0.00002	-0.00016	0.00005	0.00	0.00001	0.00001	0.00001	0.00003	0.00024	0.00005
x_3	0.00	0.00003	0.00001	0.00024	0.00002	-0.00005	0.00004	0.00	-0.00020	0.00001	-0.00007	0.00003	-0.00018	0.00006
x_4	-0.05	-0.05098	0.00006	-0.04964	0.00015	-0.04733	0.00043	-0.05	-0.05015	0.00006	-0.04761	0.00014	-0.04602	0.00042
x_5	0.00	-0.00004	0.00001	0.00037	0.00002	-0.00009	0.00005	0.00	0.00001	0.00001	0.00015	0.00003	-0.00005	0.00005
x_6	0.05	0.05027	0.00007	0.04868	0.00017	0.04719	0.00048	0.05	0.04829	0.00007	0.04852	0.00016	0.04171	0.00047
w_1	0.00	0.00051	0.00003	0.00045	0.00010	0.00050	0.00026	0.00	0.00092	0.00005	0.00109	0.00016	0.00093	0.00028
w_2	0.13	0.13947	0.00010	0.13760	0.00016	0.13619	0.00033	0.13	0.14756	0.00048	0.14869	0.00126	0.14721	0.00234
N=2000														
y_{t-1}	0.95	0.95070	0.00000	0.95034	0.00000	0.95015	0.00000	0.50	0.49044	0.00025	0.48638	0.00026	0.48469	0.00023
x_1	0.05	0.05046	0.00001	0.05039	0.00003	0.05011	0.00005	0.05	0.04907	0.00001	0.05096	0.00002	0.05250	0.00005
x_2	0.00	-0.00003	0.00000	-0.00005	0.00000	-0.00013	0.00001	0.00	0.00000	0.00000	-0.00003	0.00000	0.00001	0.00001
x_3	0.00	0.00003	0.00000	0.00014	0.00000	0.00003	0.00000	0.00	-0.00011	0.00000	-0.00003	0.00000	-0.00004	0.00000
x_4	-0.05	-0.05108	0.00001	-0.04958	0.00003	-0.04965	0.00005	-0.05	-0.05115	0.00001	-0.04866	0.00002	-0.04908	0.00005
x_5	0.00	-0.00007	0.00000	0.00025	0.00001	-0.00007	0.00001	0.00	0.00001	0.00000	0.00009	0.00000	-0.00001	0.00000
x_6	0.05	0.05037	0.00002	0.04906	0.00003	0.04932	0.00005	0.05	0.04907	0.00002	0.04903	0.00003	0.04560	0.00005
w_1	0.00	0.00011	0.00001	0.00005	0.00001	0.00023	0.00002	0.00	0.00020	0.00001	0.00026	0.00001	0.00019	0.00002
w_2	0.13	0.13350	0.00003	0.13164	0.00003	0.13103	0.00003	0.13	0.13613	0.00006	0.13498	0.00007	0.13677	0.00007

Notes:

1. The error terms are constructed using discrete distributions (see Section 4.1.).

Appendix A

Figure 1. True model vs BMA posterior estimates for various N, α, and σ_v

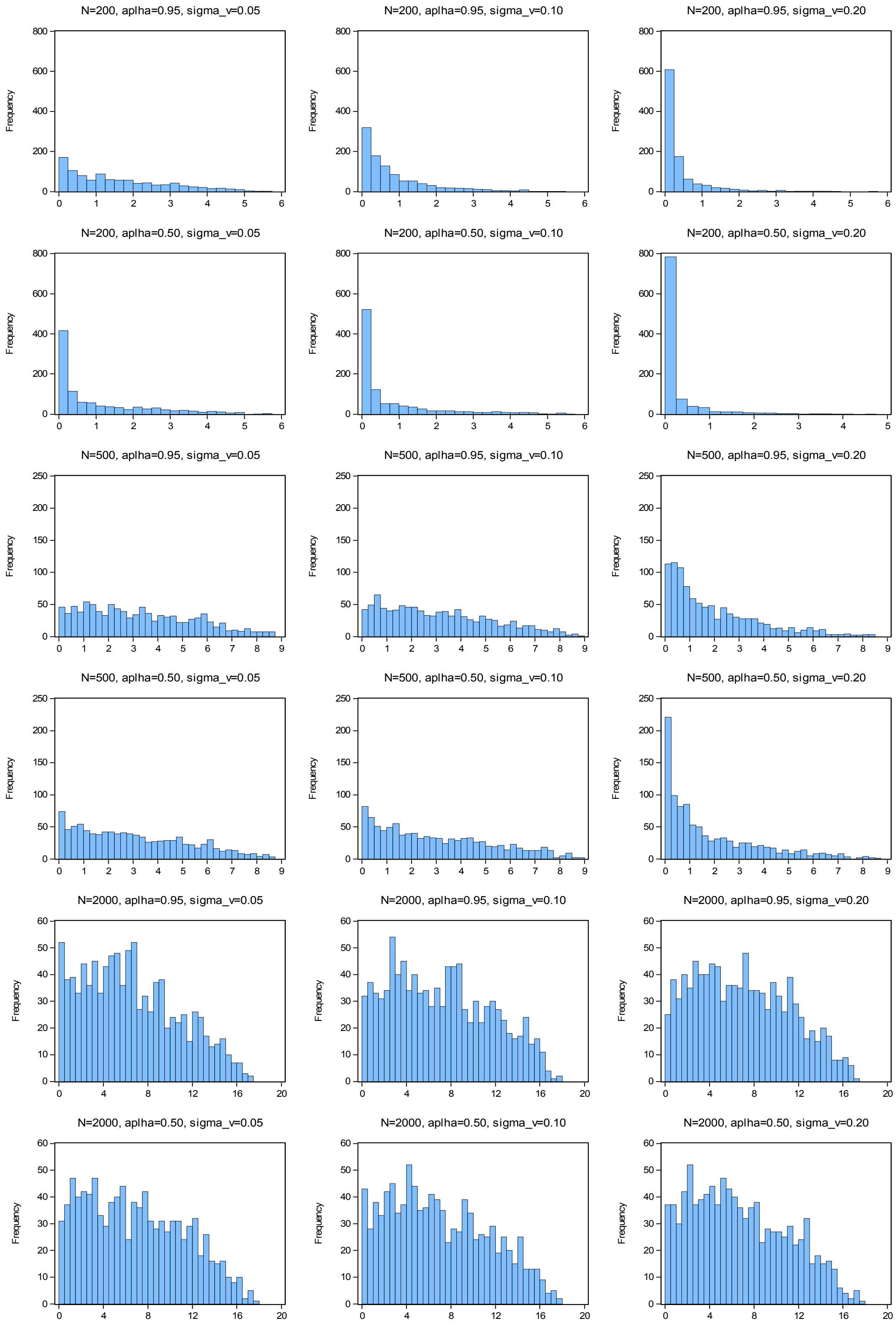


Notes:

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 2: Posterior densities for the probabilities in Table 2



N=200, aplha=0.95, sigma_v=0.05

N=200, aplha=0.95, sigma_v=0.10

N=200, aplha=0.95, sigma_v=0.20

N=200, aplha=0.50, sigma_v=0.05

N=200, aplha=0.50, sigma_v=0.10

N=200, aplha=0.50, sigma_v=0.20

N=500, aplha=0.95, sigma_v=0.05

N=500, aplha=0.95, sigma_v=0.10

N=500, aplha=0.95, sigma_v=0.20

N=500, aplha=0.50, sigma_v=0.05

N=500, aplha=0.50, sigma_v=0.10

N=500, aplha=0.50, sigma_v=0.20

N=2000, aplha=0.95, sigma_v=0.05

N=2000, aplha=0.95, sigma_v=0.10

N=2000, aplha=0.95, sigma_v=0.20

N=2000, aplha=0.50, sigma_v=0.05

N=2000, aplha=0.50, sigma_v=0.10

N=2000, aplha=0.50, sigma_v=0.20

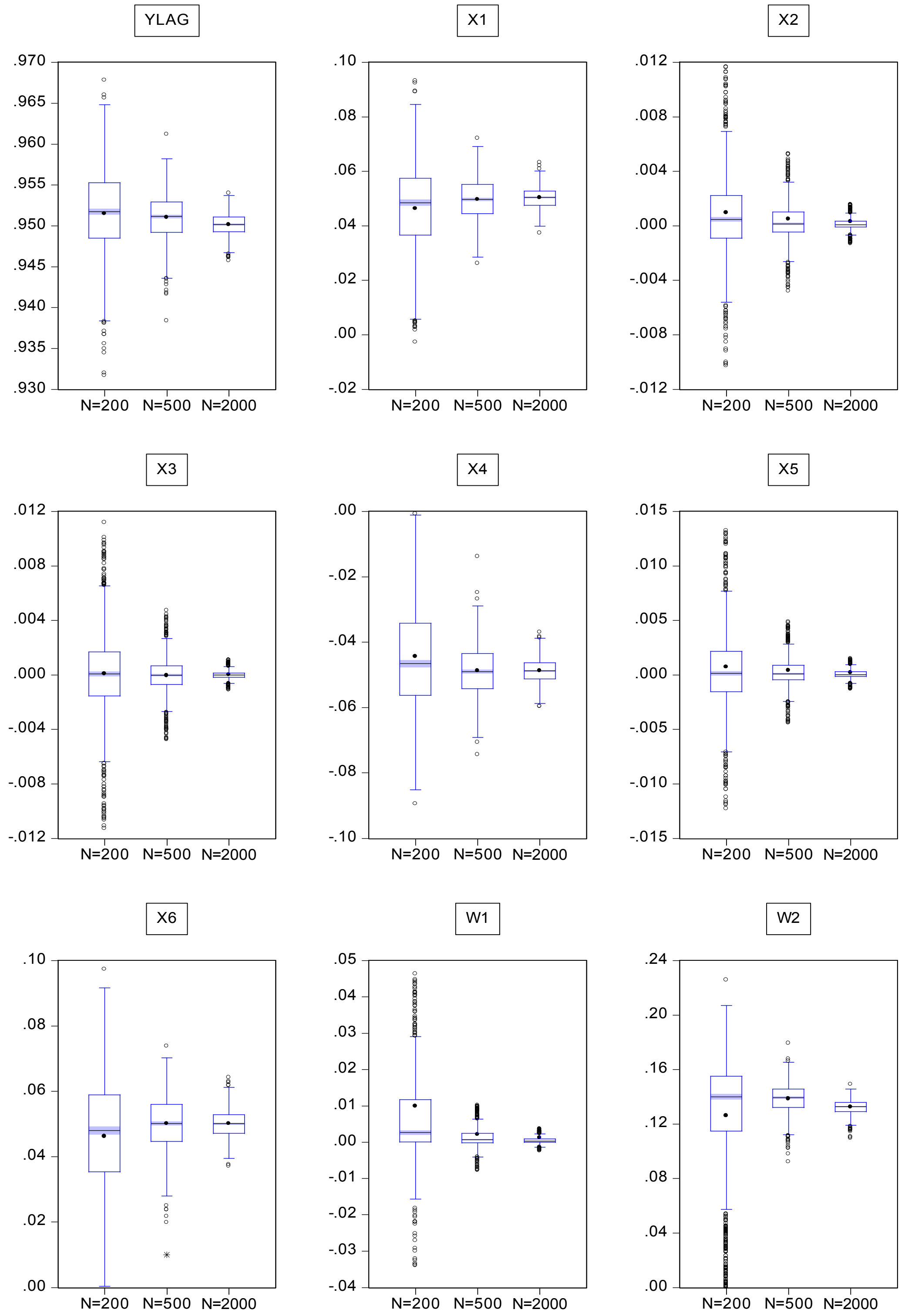
Notes:

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: Box plots for parameters in Table 5

True model vs BMA posterior estimates for various N, α, and σ_v

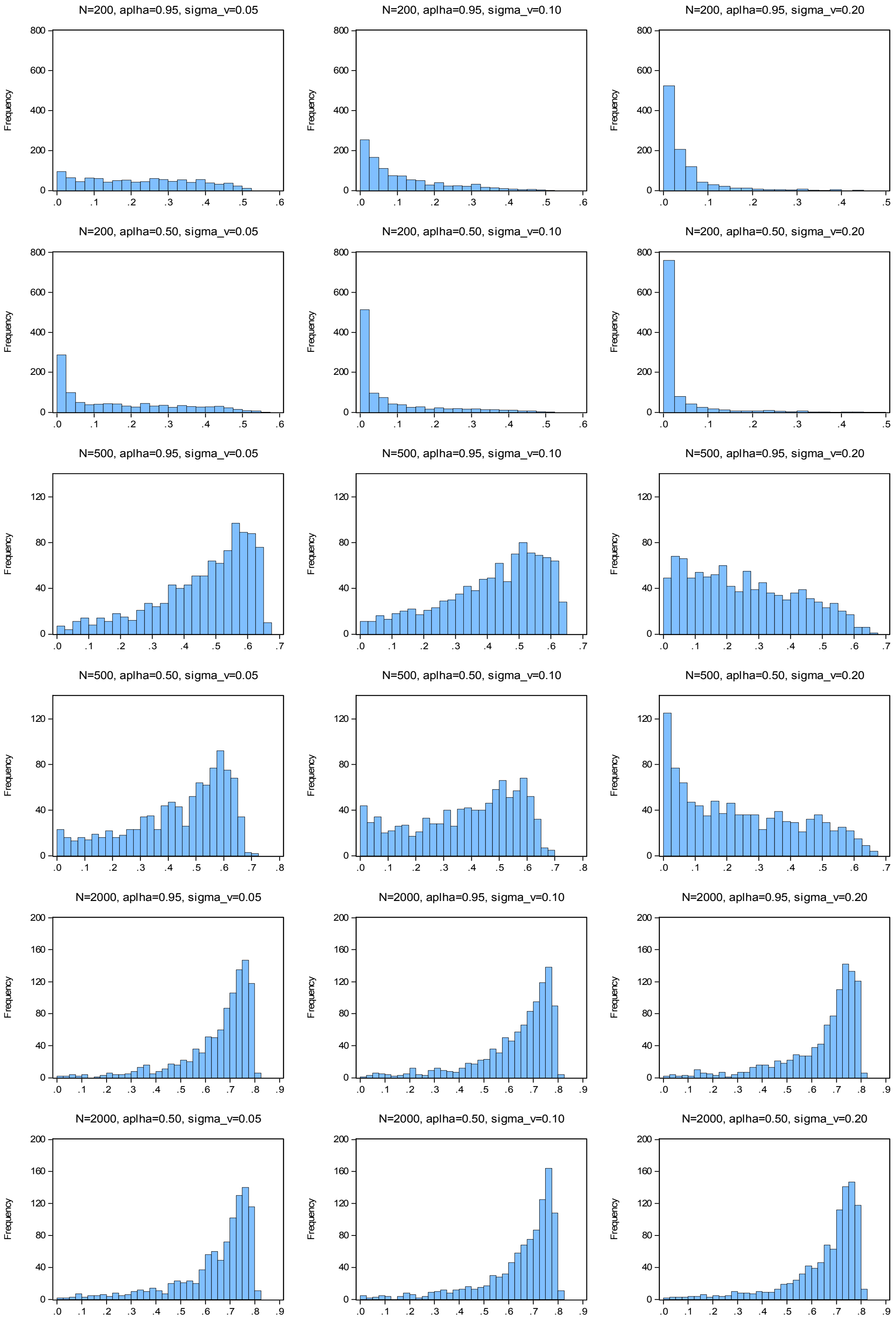


Notes:

1. For the idiosyncratic error term, $\eta_i \sim N(0,\sigma_\eta^2)$ where $\sigma_\eta^2 = 0.10$.

2. The error term is normally distributed $v_{it} \sim N(0,\sigma_v^2)$.

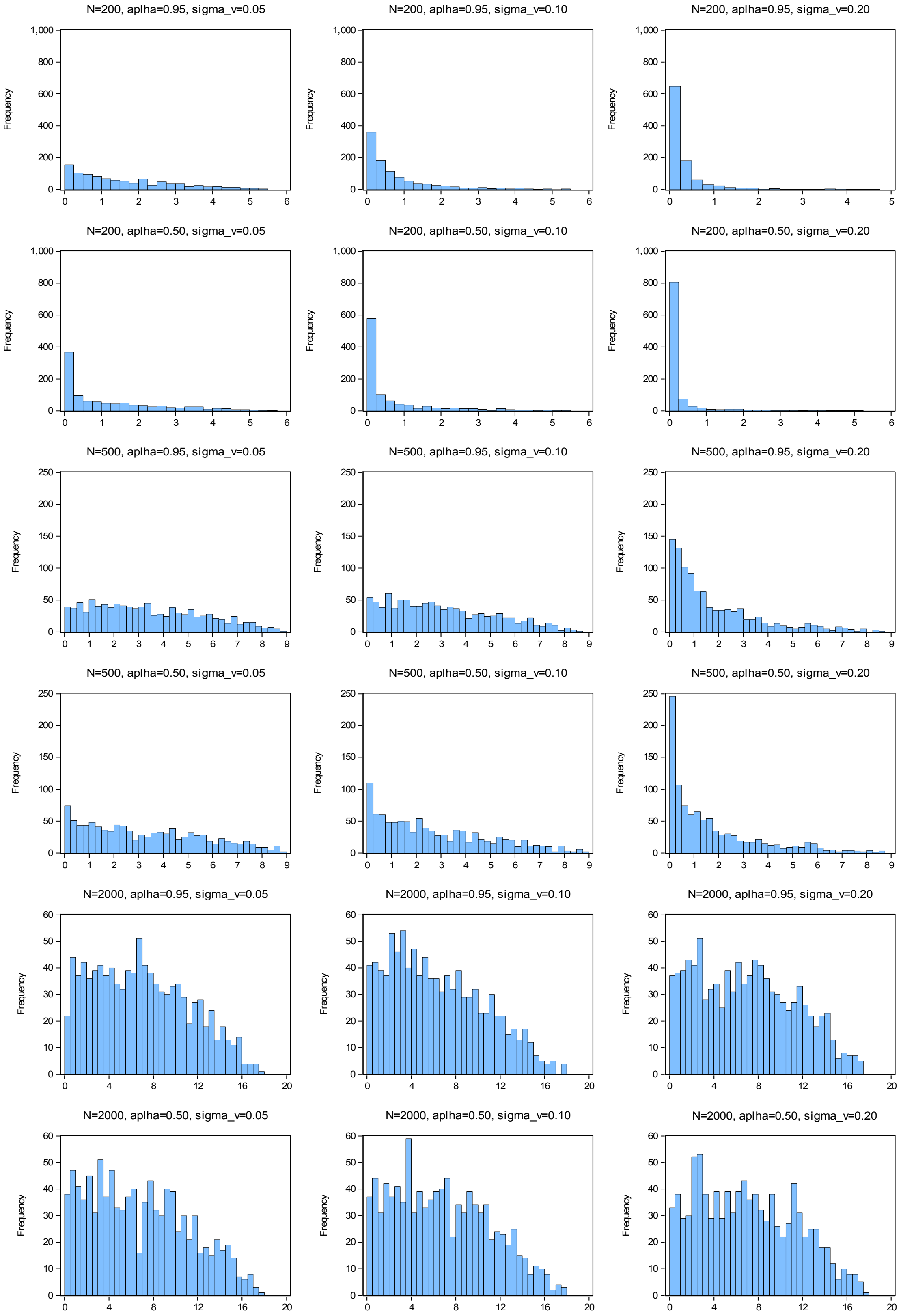
Figure 4: Posterior densities for the probabilities in Table 6



Notes:

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

Figure 5: Posterior densities for the probabilities in Table 7

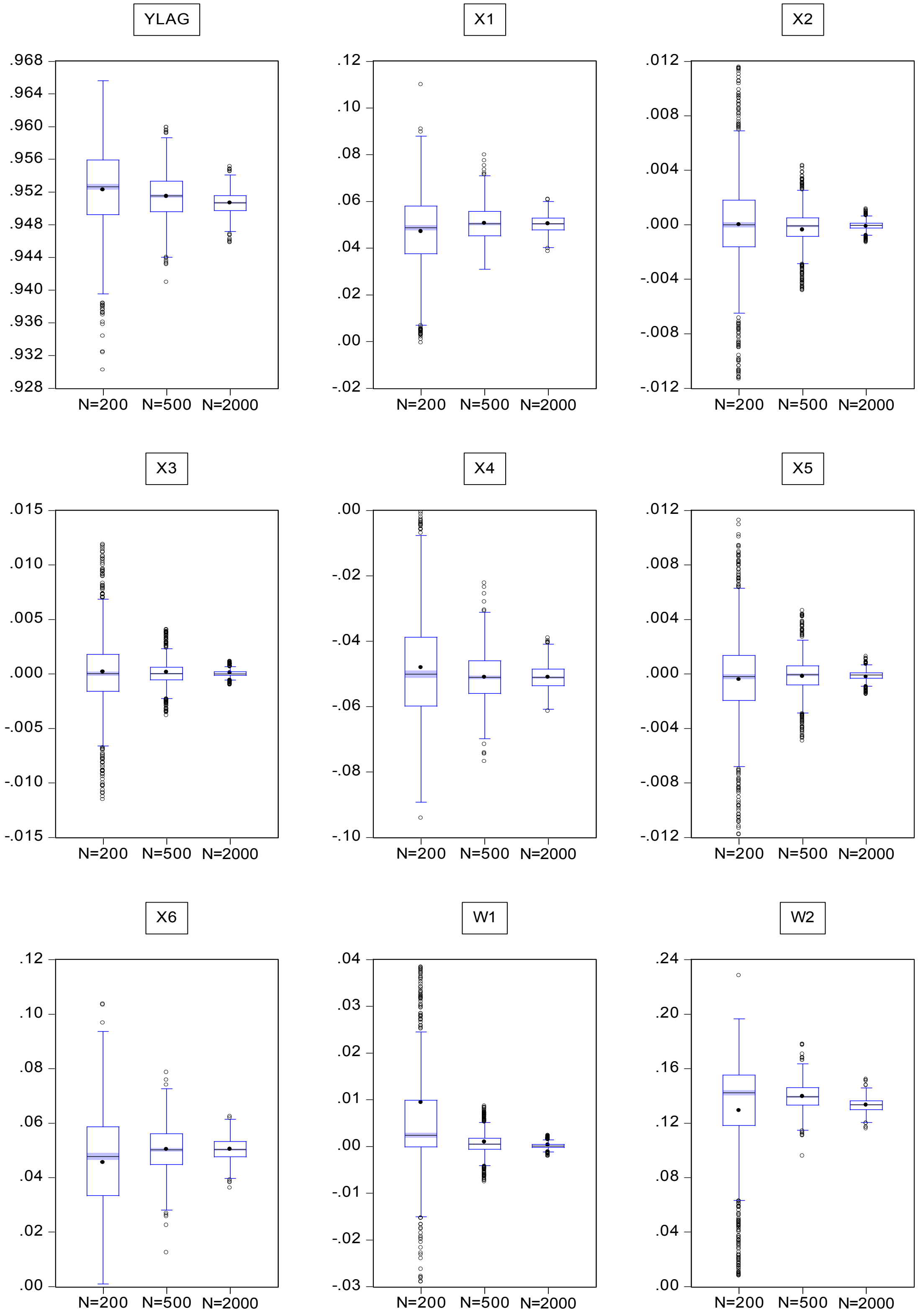


Notes:

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

Figure 6: Box plots for parameters in Table 10

True model vs BMA posterior estimates for various N, α, and σ_v



Notes:

1. The error terms are constructed using discrete distributions (see Section 4.1.).