IMF Working Paper

Limited Information Bayesian Model Averaging for Dynamic Panels with an Application to a Trade Gravity Model

Huigang Chen
Alin Mirestean
Charalambos G. Tsangarides

INTERNATIONAL MONETARY FUND

IMF Working Paper

Research Department

Limited Information Bayesian Model Averaging for Dynamic Panels with an Application to a Trade Gravity Model

Prepared by Huigang Chen, Alin Mirestean, and Charalambos G. Tsangarides[†]

Authorized for distribution by Atish R. Ghosh

October 2011

Abstract

This paper extends the Bayesian Model Averaging 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 averaging and selection. In particular, LIBMA recovers the data generating process well, with high posterior inclusion probabilities for all the relevant regressors, and parameter estimates very close to their true values. These findings suggest that our methodology is well suited for inference in short dynamic panel data models with endogenous regressors in the context of model uncertainty. We illustrate the use of LIBMA in an application to the estimation of a dynamic gravity model for bilateral trade.

Contents

I.   Introduction

Model uncertainty is an issue encountered often in the econometric study of socioeconomic phenomena. Typically, model uncertainty arises because the lack of clear theoretical guidance and trade-offs in the choice of regressors result in a broad number of possible and testable specifications, a phenomenon labeled as "open-endedness" of economic theories by Brock and Durlauf (2001). Since several modes may be plausible given the data, inferences about parameters of interest may be fragile and may even result in contradictory conclusions. Typical attempts to deal with model uncertainty include down-playing its importance (Temple (2000)) and engaging in unsystematic searches of possible model specifications.

A growing number of researchers are turning to the Bayesian methods in order to deal with the problem of model uncertainty. Bayesian model selection attempts to find the data generating process by choosing the most probable model in the universe of models, that is, the model for which the posterior model probability is the largest (see George (1999), and Chipman, George and McCulloch (2001)). The posterior model probability provides a representation of the model uncertainty conditional on the information contained in the observed data, and inference pertaining to a quantity of interest is conditional on the "selected model" model. However, this may underestimate the uncertainty related to that quantity of interest since uncertainty across different models is not accounted for. Bayesian Model Averaging (BMA)—initially proposed by Leamer (1978)—fully incorporates model uncertainty by basing inferences of a quantity of interest on an average of the posterior distributions under each model, using posterior model probabilities as weights. Contributions to the BMA literature include those of Moulton (1991), Madigan and Raftery (1994), Kass and Raftery (1995), Raftery, Madigan and Hoeting (1997), and Fernández, Ley and Steel (2001b). The BMA framework has been applied in various areas of social sciences, including economics.[1]

Despite the increasing interest in BMA to address model uncertainty, most of the work so far uses static models and cross section analysis with data averaged over the time dimension, effectively ignoring dynamic relationships among variables. Moreover, to the best of our knowledge, none of the models allow for the inclusion of endogenous variables, namely regressors that are correlated with the disturbances.[2] In this paper, we first propose a limited information 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 label our methodology "Limited Information Bayesian Model Averaging" (LIBMA). Then, we evaluate the performance of LIBMA relative to both Bayesian model selection and Bayesian Model Averaging using Monte Carlo simulations. Finally, we present an application of LIBMA to the estimation of a dynamic panel data gravity model for bilateral trade, which illustrates its usefulness in practice.

---

[1] In the growth empirics literature, work includes Brock and Durlauf (2001), Fernández, Ley and Steel (2001a), Sala-i-Martin, Doppelhofer and Miller (2004), and Temple and Malik (2009). Other applications include biology (Yeung, Bumgarner, and Raftery (2005)), ecology (Wintle et al. (2003)), public health (Morales et al. (2006)), and toxicology (Koop and Tole (2004)).

[2] Moral–Benito (2010) considers a panel data model where the lagged dependent variable (only) is correlated with the individual effects but not correlated with the idiosynchratic error term.

Specifically, we propose a method for constructing the model likelihoods and posteriors based only on information elicited from moment conditions, with no specific distributional assumptions. The limited information likelihood we build is a "true" likelihood, derived by taking advantage of the linear structure of the model, as well as the asymptotic properties of the Generalized Method of Moments (GMM) estimator via the central limit theorem. Therefore, for the purpose of Bayesian inference, our limited information likelihood can be used, in principle, with any prior. In this paper, however, we only consider a likelihood dependent, unit information prior (see Kass and Wasserman (1995)) which provides a posterior with a simple Bayesian Information Criterion (BIC)-like form with a good approximation, which also reduces the computational burden in obtaining model posteriors. In addition, we derive the marginal likelihood using standard Bayesian procedures, that is, integrating out the parameters in the likelihood function, which satisfies large sample properties automatically. Our approach differs from earlier work by Kim (2002), Tsangarides (2004), and Hong and Preston (2008) that approximates the model marginal likelihood by quasi likelihood functions whose specifications are justified only through their large sample properties. Finally, our approach is similar in spirit to the work of Schennach (2005) and Ragusa (2008) in the sense that the likelihood is constructed using a Bayesian procedure, but it is simpler in construction.

Section 2 introduces the concept of model uncertainty in the Bayesian context and then reviews model selection and model averaging. Section 3 develops the theoretical framework of the LIBMA methodology in the context of dynamic panels with endogenous regressors. Section 4 discusses the proposed simulation experiment and presents the results. Section 5 presents an application of LIBMA to the estimation of a dynamic gravity model for bilateral trade. The final section concludes.

II.   Model Uncertainty in the Bayesian Context

For completeness, we begin by reviewing briefly the basic theory of uncertainty in the Bayesian context. Excellent discussions include Hoeting, Madigan, Raftery and Volinsky (1999), and Chipman, George and McCulloch (2001).

A.   Model Selection and Hypothesis Testing

Suppose there is a universe of K possible explanatory variables $x_1, x_2, \ldots, x_K$. Let Z be the design matrix of all possible explanatory variables. For a given model $M \subset U = \{1, 2, \ldots, K\}$, we consider the standard linear regression model

$$Y = Z_M \theta_M + u \tag{1}$$

where Y is the variable of interest, $\theta = (\theta_1, \ldots, \theta_K)'$ is a vector of parameters to be estimated, and u is the error term. We use M to represent that only the part that corresponds to the variables whose index belongs to M is selected.

Given the universe of K possible explanatory variables, a set of $\mathcal{K} = 2^K$ models $\mathcal{M} = (M_1, ..., M_{\mathcal{K}})$ are under consideration. In the spirit of Bayesian inference, priors

$p(\theta|M_j)$ for the parameters of each model, and a prior $p(M_j)$ for each model in the model space $\mathcal{M}$ can be specified.

Model selection seeks to find the model M_j in $\mathcal{M} = (M_1, ..., M_{\mathcal{K}})$ 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

$$p(M_j|D) = \frac{p(D|M_j)p(M_j)}{\sum_{l=1}^{\mathcal{K}} p(D|M_l)p(M_l)} \tag{2}$$

where

$$p(D|M_j) = \int p(D|\theta_j, M_j)p(\theta_j|M_j)d\theta_j \tag{3}$$

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) 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 (as also in Fernández, Ley and Steel (2001b)), 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 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

$$p(\Gamma|D) = \sum_{j=1}^{\mathcal{K}} p(\Gamma|D, M_j)p(M_j|D) \tag{4}$$

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 $(M_1, ..., M_{\mathcal{K}})$, where the weights are the posterior model probabilities $p(M_j|D)$. From the linear regression model (1), BMA allows the computation of the inclusion probability for every possible explanatory variable.

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

where

$$I(Z_i|M_j) = \begin{cases} 1 & if \quad Z_i \in M_j \\ 0 & if \quad Z_i \notin M_j \end{cases}.$$

Using (4) one can compute the posterior mean for parameters θ_l as follows

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

Implementation of BMA presents a number of challenges, including the evaluation of the marginal probability in (3), the large number of possible models, and the specification of prior model probabilities $p(M_j)$, and the parameters, $p(\theta|M_i)$.

C. Choice of Priors

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

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

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,

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

In this case, determining the prior $p(\theta, \zeta|M_i)$ becomes an important issue.[3]

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 the Normal-Gamma distribution, which has the benefit of rendering a closed-form posterior.[4] 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 $\{\theta_0, V, \gamma, \lambda\}$, choosing specific values for them is often avoided. 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 $p(\sigma_u) \propto \sigma_u^{-1}$. 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 Zellner's g-prior

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

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

The prior used in this paper is another choice of prior for Bayes factors when data is an i.i.d. sequence of observations, namely, the so-called unit information prior. Suppose we

---

[3] 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.

[4] For a more detailed discussion see Kass and Wasserman (1995).

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.

For the specification of model priors $p(M_j)$, several options exist. Model priors may reflect the researcher's view about the number of regressors that should be included, with a penalty that increases proportionally with the number of regressors included in the model. Such a prior model probability structure which reflects the researcher's prior about the size of the model, was initially proposed by Mitchell and Beauchamp (1988) and was used by Sala-i-Martin, Doppelhofer and Miller (2004).[5] We choose to specify model priors using an alternative approach used by Fernández, Ley and Steel (2001b). We assume a Uniform distribution over the model space which implies that we have no prior preference for a specific model, so that $p(M_1) = p(M_2) = ... = p(M_{\mathcal{K}}) = \frac{1}{\mathcal{K}}$.[6]

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

Consider the case where a researcher is faced with model uncertainty when trying to estimate a dynamic model for panel data. 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 $\{\{1\}, X, W\}$ is a partition of U. Therefore, for a given model $M_j \subset U$, (1) becomes

$$
\begin{aligned}
y_{it} &= \left(\begin{array}{ccc} y_{i,t-1} & x_{it} & w_{it} \end{array} \right)_{M_j} \left(\begin{array}{ccc} \alpha & \theta_x & \theta_w \end{array} \right)'_{M_j} + u_{it} \\
u_{it} &= \eta_i + v_{it} \\
|\alpha| &< 1; \ i = 1, 2, ..., N; \ t = 1, 2, ..., T
\end{aligned}
\tag{5}
$$

---

[5] Specifically, the prior probability for model M_j is $p(M_j) = \left(\frac{\overline{k}}{K} \right)^{k_j} \left(1 - \frac{\overline{k}}{K} \right)^{K-k_j}$. The prior odds ratio is $\frac{p(M_j)}{p(M_l)} = \left(\frac{\overline{k}}{K} \right)^{k_j-k_l} \left(1 - \frac{\overline{k}}{K} \right)^{k_l-k_j}$ 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.

[6] Comparing diffent prior structures considered in the BMA literature, Eicher, Papageorgiou, and Raftery (2010) find that the combination of Unit Information Prior over the parameter space and the uniform model prior–the priors used in this paper–outperform any combination of priors previously in terms of cross-validated predictive performance.

where y_{it}, x_{it}, and w_{it} are observed variables, η_i is the unobserved individual effect, and 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. We assume that $E(v_{it}) = 0$ and that v_{it}'s are not serially correlated, x_{it} is a $1 \times m$ vector of exogenous variables, and 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

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

Similarly, for any endogenous variable we have

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

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 (5) is to use the system GMM framework (Arellano and Bond (1991), Arellano and Bover (1995), and Blundell and Bond (1998)). This implies constructing the instrument set and moment conditions for the "level equation" (5) and combining them with the moment conditions using the instruments corresponding to the "first difference" equation. The first difference equation corresponding to model (5) is given by

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

One assumption required for the first difference equation is that the initial value of y, y_{i0}, is predetermined, that is, $E(y_{i0} v_{is}) = 0$ 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 $E(y_{i,t-2} \Delta v_{it}) = 0$ for $t = 2, 3, ..., T$. Moreover, $y_{i,t-3}$ is 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

$$E(y_{i,t-s} \Delta v_{it}) = 0, \ t = 2, 3, ..., T; \ s = 2, 3, ..., t; \ \text{for } T \geq 2, \ i = 1, 2, ..., N. \tag{6}$$

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. That gives us additional moment conditions

$$E(x_{it}^l \Delta v_{it}) = 0, \ t = 2, 3, ..., T; \ l = 1, ..., m; \ i = 1, 2, ..., N. \tag{7}$$

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

$$E(w^l_{i,t-s}\Delta v_{it}) = 0, \ t = 3, 4, ..., T; \ s = 2, ...t - 1;$$
$$\text{for } T \geq 3, \ l = 1, 2, ..., q; \ i = 1, ..., N. \tag{8}$$

Table A summarizes the moment conditions that could be used for the first difference 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 first difference 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.

Returning to the level equation (5), 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

$$E(\Delta y_{i,t-1}u_{it}) = 0, \ t = 2, 3, ..., T. \tag{9}$$

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

$$E(\Delta w^l_{i,t-1}u_{it}) = 0, \ t = 3, 4, ..., T, \ l = 1, 2, ..., q. \tag{10}$$

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 so another set of moment conditions can be used

$$E(\Delta x^l_{it}u_{it}) = 0, \ t = 2, 3, ..., T, \ l = 1, 2, ..., m. \tag{11}$$

Table B summarizes the moment conditions for the level equation. There are $(T-1)$ 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.

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

As shown by Ahn and Schmidt (1995), an additional $(T-1)$ 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,

$$E(y_{i,t} u_{i,t} - y_{i,t-1} u_{i,t-1}) = 0, \ t = 2, 3, ..., T; \ i = 1, ..., N. \tag{12}$$

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

$$\sum_{t=2}^{T} E(x_{it}^l \Delta v_{it}) + \sum_{t=2}^{T} E(\Delta x_{it}^l u_{it}) = 0, \quad l = 1, ..., m; \ \ i = 1, 2, ..., N. \tag{13}$$

All the above moment conditions can be succinctly written in matrix form

$$E\left[G_i' U_i\right] = 0. \tag{14}$$

Definitions of matrices G_i and U_i are presented in Appendix A.

Based on the moment conditions (14) we propose a limited information criterion that can be used in Bayesian model selection and averaging. The next section provides details on how to construct this criterion.

C.   The Limited Information Criterion

As pointed out in Section 2, evaluating the Bayes factors needed for hypothesis testing and Bayesian model selection or model averaging requires calculating the marginal likelihood

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

Since GMM is chosen to estimate the parameters of the model, the assumptions made so far do not give a fully specified parametric likelihood $p\left(D|\theta, M_i\right)$. 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 a lot of attention in the literature lately. Several approaches have been used to derive or estimate non-parametric likelihood functions, including Back and Brown (1993), Kim (2002), Schennach (2005), Hong and Preston (2008), and Ragusa (2008). We propose a method of constructing the model likelihoods and posteriors based only on the information elicited from the moment conditions (14). 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.

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. $E\left[g\left(\xi_i, \theta\right)\right]$ exists and is finite for every $\theta \in \Theta$; and

3. $E\left[g\left(\xi_i, \theta\right)\right]$ is continuous on θ.

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

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

For model (5), the moment conditions for individual i discussed in the previous section can be written in the following form

$$g\left(\xi_i, \theta\right) = G_i'\left(\widetilde{y}_i - \widetilde{z}_i\theta\right),$$

where $\xi_i = \{\widetilde{y}_i, \widetilde{z}_i\}$, $\widetilde{z}_i = \left(\begin{array}{ccc} \widetilde{y}_{i,-1} & \widetilde{x}_i & \widetilde{w}_i \end{array}\right)$, $\theta = \left(\begin{array}{ccc} \alpha & \theta_x & \theta_w \end{array}\right)'$, while G_i is the matrix defined in (14). The vectors $\widetilde{y}_i$ and $\widetilde{y}_{i,-1}$ for the dependent variable and the lagged dependent variable, respectively, are defined as follows

$$\begin{aligned}
\widetilde{y}_i &= \left(\begin{array}{ccccccc} y_{i1} & y_{i2} & \cdots & y_{iT} & \Delta y_{i2} & \Delta y_{i3} & \cdots & \Delta y_{iT} \end{array}\right)' \\
\widetilde{y}_{i,-1} &= \left(\begin{array}{ccccccc} y_{i0} & y_{i1} & \cdots & y_{i,T-1} & \Delta y_{i1} & \Delta y_{i2} & \cdots & \Delta y_{i,T-1} \end{array}\right)'.
\end{aligned}$$

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

$$\widetilde{x}_i = \left(\begin{array}{ccccc}
x_{i1}^1 & x_{i1}^2 & x_{i1}^3 & \cdots & x_{i1}^m \\
\vdots & \vdots & \vdots & \cdots & \vdots \\
x_{iT}^1 & x_{iT}^2 & x_{iT}^3 & \cdots & x_{iT}^m \\
\Delta x_{i2}^1 & \Delta x_{i2}^2 & \Delta x_{i2}^3 & \cdots & \Delta x_{i2}^m \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\Delta x_{iT}^1 & \Delta x_{iT}^2 & \Delta x_{iT}^3 & \cdots & \Delta x_{iT}^m
\end{array}\right),$$

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

$$\widetilde{w}_i = \left(\begin{array}{ccccc}
w_{i1}^1 & w_{i1}^2 & w_{i1}^3 & \cdots & w_{i1}^q \\
\vdots & \vdots & \vdots & \cdots & \vdots \\
w_{iT}^1 & w_{iT}^2 & w_{iT}^3 & \cdots & w_{iT}^q \\
\Delta w_{i2}^1 & \Delta w_{i2}^2 & \Delta w_{i2}^3 & \cdots & \Delta w_{i2}^q \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\Delta w_{iT}^1 & \Delta w_{iT}^2 & \Delta w_{iT}^3 & \cdots & \Delta w_{iT}^q
\end{array}\right).$$

Therefore, $\widehat{g}_N(\theta_0) = N^{-1} \sum_{i=1}^{N} G_i'\widetilde{y}_i - N^{-1} \sum_{i=1}^{N} G'\widetilde{z}_i \theta_0$.

Since $\{y_{i0}, x_{i\cdot}, z_{i\cdot}, u_{i\cdot}\}$ is assumed to be a strictly stationary and ergodic process with finite second moment, $E[\partial g(\xi_i, \theta)/\partial\theta] = -E[\widetilde{z}_i'G_i]$ is finite and has full rank by the choice of moment conditions. Therefore, by standard argument (see Hansen 1982), $E[g(\xi_i, \theta)]$ is continuous on θ. In addition, $\omega_i \equiv g(\xi_i, \theta_0)$ is stationary and independent. It satisfies that $E[\omega_i] = 0$, $E[\omega_i\omega_i']$ exists and is finite positive definite. This ensures that $\lim_{N\to\infty} Var[N^{1/2}\widehat{g}_N(\theta_0)]$ exists, is finite, and is positive definite (Hansen 1982). Therefore, Lemma 1 can be applied to our dynamic panel data model.

By Lemma 1, the likelihood for θ can be written as

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

and the model likelihood can be expressed as

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

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

$$\int_\Theta p\left(N^{-1}\sum_{i=1}^{N} G_i'\widetilde{y}_i\,\big|\,\theta\right) p(\theta)\, d\theta$$

$$\propto \exp\left(\begin{array}{c} -\frac{1}{2}N\widehat{g}_N'\left(\widehat{\theta}_0\right)S^{-1}\left(\widehat{\theta}_0\right)\widehat{g}_N\left(\widehat{\theta}_0\right) + \log p\left(\widehat{\theta}_0\right) \\ +\frac{K}{2}\log 2\pi - \frac{1}{2}\log\det\frac{\partial^2}{\partial\theta^2}\left(\frac{1}{2}N\widehat{g}_N'\left(\widehat{\theta}_0\right)S^{-1}\left(\widehat{\theta}_0\right)\widehat{g}_N\left(\widehat{\theta}_0\right)\right) \end{array} \right),$$

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

$$\int_\Theta p\left(N^{-1}\sum_{i=1}^{N} G_i'\widetilde{y}_i\,\big|\,\theta\right) p(\theta)\, d\theta$$

$$\propto \exp\left(-\frac{1}{2}N\widehat{g}_N'\left(\widehat{\theta}_0\right)S^{-1}\left(\widehat{\theta}_0\right)\widehat{g}_N\left(\widehat{\theta}_0\right) - \frac{K}{2}\log N\right), \quad (15)$$

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 $\partial^2\left(\frac{1}{2}\widehat{g}_N'S^{-1}\widehat{g}_N\right)/\partial\theta^2|_{\theta=\widehat{\theta}_0}$ as its variance-covariance matrix. That is, the prior distribution for θ, $p(\theta)$, is given by $N\left(\widehat{\theta}_0, \partial^2\left(\frac{1}{2}\widehat{g}_N'S^{-1}\widehat{g}_N\right)/\partial\theta^2|_{\theta=\widehat{\theta}_0}\right)$.

The unit information prior is a type of weakly informative prior based on observed data. The precision of $\widehat{\theta}_0$, or its inverse variance, from the GMM estimator is $\partial^2\left(\frac{1}{2}N\widehat{g}_N'S^{-1}\widehat{g}_N\right)/\partial\theta^2|_{\theta=\widehat{\theta}_0}$. This can be considered as the amount of information from the entire observation, and can be taken as $\partial^2\left(\frac{1}{2}\widehat{g}_N'S^{-1}\widehat{g}_N\right)/\partial\theta^2|_{\theta=\widehat{\theta}_0}$, where $\widehat{\theta}_0$ is used as

the mean of the prior (as suggested by Kass and Wasserman (1995)). The Gaussian distribution is the least "informative" distribution (maximum entropy distribution) with given mean and variance. So, the use of unit information prior can be thought of as the prior distribution from a person who has unbiased but weak information about the coefficients.[7]

For a given model M_j for which θ has k_j elements different from zero and with the estimate denoted by $\widehat{\theta}_{0,j}$, the model likelihood (15) becomes

$$\int_\Theta p\left(N^{-1}\sum_{i=1}^N G_i'\widetilde{y}_i \,|\theta, M_j\right) p(\theta)\, d\theta$$

$$\propto \exp\left(-\frac{1}{2}N\widehat{g}_N'\left(\widehat{\theta}_{0,j}\right) S^{-1}\left(\widehat{\theta}_{0,j}\right)\widehat{g}_N\left(\widehat{\theta}_{0,j}\right) - \frac{k_j}{2}\log N\right). \quad (16)$$

Then, the moment conditions (14) associated with model M_j can be written as

$$E\left[G_i'\left(\widetilde{y}_i - (\widetilde{z}_i)_{M_j}(\theta_0)_{M_j}\right)\right] = 0,$$

where the subscript M_j indicates that only entries corresponding to model M_j are taken. 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
$\widehat{g}_N\left(\widehat{\theta}_{0,j}\right) = N^{-1}\sum_{i=1}^N G_i'\left(\widetilde{y}_i - (\widetilde{z})_{M_j}\widehat{\theta}_{0,j}\right)$. 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] which allows valid comparisons of posterior probabilities, based on the principle of Bayesian factor analysis. Therefore, by using (16), we can compute the posterior odds ratio of two models M_1 and M_2 by

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

$$= \frac{p(M_1)}{p(M_2)}\exp\left(\begin{array}{c} -\frac{1}{2}N\widehat{g}_N'\left(\widehat{\theta}_{0,1}\right) S^{-1}\left(\widehat{\theta}_{0,1}\right)\widehat{g}_N\left(\widehat{\theta}_{0,1}\right) \\ +\frac{1}{2}\widehat{g}_N'\left(\widehat{\theta}_{0,2}\right) S^{-1}\left(\widehat{\theta}_{0,2}\right)\widehat{g}_N\left(\widehat{\theta}_{0,2}\right) \\ -\left(\frac{k_1-k_2}{2}\log N\right) \end{array}\right), \quad (17)$$

which has the same form of BIC as fully specified models. We use iterative GMM estimation with moment conditions $E\left[G_i'\left(\widetilde{y}_i - (\widetilde{z}_i)_{M_j}\theta_{0,j}\right)\right] = 0$ 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 (17). As discussed in Section 2, we assume a unit information prior for the parameters and a Uniform distribution over the model space, essentially implying that there is no preference for a specific model, so $p(M_1) = p(M_2) = ... = p(M_\mathcal{K}) = \frac{1}{\mathcal{K}}$.

---

[7] Note that the above argument is general and applies to any model being estimated by GMM, regardless of its particular structure, i.e. the inclusion of the lag dependent, predetermined or endogenous variables.

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

IV.   Monte Carlo Simulation and Results

This section describes 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 9 variables, namely, 6 exogenous variables, 2 endogenous variables and the lagged dependent variable. Throughout our simulations we keep the number of periods constant at $T = 4$, and vary the number of individuals, N.

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

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

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

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

where I_2 is the two dimensional identity matrix.

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

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

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

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

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

where $\theta_x = \begin{pmatrix} 0.05 & 0 & 0 & -0.05 & 0 & 0.05 \end{pmatrix}'$, $\theta_w = \begin{pmatrix} 0 & 0.13 \end{pmatrix}'$, $w_{i0} = \begin{pmatrix} w_{i0}^1 & w_{i0}^2 \end{pmatrix}$, 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 $t = 1, 2, ..., T$ the data generating process is given by

$$y_{it} = \alpha y_{i,t-1} + \theta_x x_{it} + \theta_w w_{it} + \eta_i + v_{it}$$
$$\text{with } v_{it} \sim N\left(0, \sigma_v^2\right) \text{ and } \eta_i \sim N\left(0, \sigma_\eta^2\right).^9$$

Further, 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. The distribution of the random variable v_{it}, is obtained as follows. First, we 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 $\omega_k \sim Exponential\,(1)$. 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 the 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}.

B.   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 present results in the form of medians, means, variances and quartiles.[10] We consider sample sizes of $N = 200$, 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 $\sigma_\eta^2 = 0.10$. As discussed earlier, we examine the robustness of our results by creating a second set of simulations where the assumption of normality for v_{it} is relaxed.

1.   Model Selection

The posterior model probability is a key indicator of performance in the Bayesian framework. 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

---

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

[10] Several applications in the BMA literature suggest that Markov Chain Monte Carlo methods often give good approximations of the model posteriors when the model space is too large to manage computationally.

values ranging from 0.633 to 0.655. In addition, median posterior model probabilities become slightly higher than the means, ranging from 0.690 to 0.705 for $N = 2000$, and 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.[11]

Equation (2) shows 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, this would have an effect on the absolute value of the estimated posterior model probability. Taking this into account, we compute an additional (relative) measure, that is 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 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 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 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.

2. Model Averaging

In several cases, 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. One measure that we report for our experiments is the inclusion probability for each variable considered. The inclusion probability for a given variable is a measure of how much the data favors the inclusion of the variable, and is calculated 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, Table 4).[12] Given the assumptions made relative to the model priors, the prior probability of inclusion for each variable is the same and equal to 0.50. 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

---

[11] Figures 1, 2, 4, and 5 show density plots for the posteriors in Tables 1, 2, 6, and 7.
[12] A value of 1(0) in column 2 indicates that the true model contains (excludes) that variable.

the sample size with the upper bound being less than 0.07 for all the cases considered when $N = 2000$. Importantly, it is interesting to see that even in cases where the recovery rate of the true model is poor (for example, 12 percent for the case in which $N = 200, \alpha = 0.95$, and $\sigma_v^2 = 0.20$), 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.[13] 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, we present the box plots for the parameter estimates of Table 5, for the case of $\alpha = 0.95$ and $\sigma_v = 0.1$. As the sample increases, the variance of the distribution decreases and the median converges to the true value. In addition to 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.

Overall, while model selection properties are also desired, the strength of our methodology is given by its performance in the context of Bayesian Model Averaging.

3. Robustness Checks Using non-Gaussian Errors

We relax the normality assumption for the error term v_{it} and check the robustness of our results. 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, as 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. Results in Table 7 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

---

[13] 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.

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.

V. LIBMA Application to a Dynamic Trade Gravity Model

A. Background

Following the work of Rose (2000), there has been considerable interest in investigating the determinants of bilateral trade using gravity models. Augmented versions of trade gravity models have been used to examine the effects of exchange rate regimes, exchange rate volatility, and free trade agreements (FTAs) on bilateral trade.[14]

While relevant studies almost unanimously find that exchange rate regimes with lower uncertainty and transaction costs—namely, conventional pegs and currency unions—are significantly more pro-trade than flexible regimes, there is considerable uncertainty about the size of this effect, given the potential reverse causality between trade and currency unions. Another point of debate is whether FTAs are trade creating (as they could create trade that would not have existed otherwise from a more efficient producer of the product) or trade diverting (as trade may be diverted away from a more efficient supplier outside the FTA towards a less efficient supplier within the FTA), with several studies finding different results depending on the set of regressors used.[15] Finally, the majority of gravity models are static, ignoring potential trade persistence effects which may arise from trade

---

[14] Anderson (1979) and Bergstrand (1985) provide theoretical foundations for the use of gravity models in bilateral trade. Empirical applications include Frankel and Rose (2002), Klein and Shambaugh (2006), and Santos Silva and Tenreyro (2006).

[15] See Ghosh and Yamarik (2004) and Baxter and Kouparitsas (2006). Eicher, Henn, and Papageorgiou (2011) use Raftery's BMA in a static gravity model (without accounting for potential endogeneity and dynamics).

inertia due to sunk costs associated with distribution and service networks, or habit formation due to consumers getting accustomed to foreign products.[16]

These observations suggest that the investigation of bilateral trade determinants is complicated by the fact that several determinants are potentially endogenous, and the considerable uncertainty about the set of explanatory variables that should be included in the gravity model specification. Model uncertainty combined with the use of a dynamic specification with several endogenous regressors suggest that our LIBMA is well suited for the estimation of a dynamic trade gravity model.

B. Model and Estimation

In its simple form, the static gravity trade equation can be expressed as

$$X_{ij} = Y_i Y_j \left(\frac{T_{ij}}{P_i P_j} \right)^{1-\sigma} \tag{18}$$

where X_{ij} represents average trade from country i to j; Y_i and Y_j are total domestic outputs in country i and j, respectively; P_i and P_j are the overall price indices in country i and j, respectively; T_{ij} are iceberg trading costs; and σ is the elasticity of substitution between products ($\sigma > 1$). Traditionally, T_{ij} in equation (18) includes transportation costs that are proxied by geographical attributes (such as bilateral distance, access to sea, and contiguity). In recent years, other factors that may affect trade costs, for example, common language, historical ties, FTAs, tariffs, and non-tariff barriers have also been included. To the extent that exchange rate policy choices influence currency conversion costs, exchange rate volatility as well as uncertainty, trading costs would also depend on the exchange rate regime in place, making its inclusion in T_{ij} appropriate.

Qureshi and Tsangarides (2010) examine the trade effects of exchange rate regimes by augmenting the traditional gravity equation and including variables for currency unions, conventional pegs, and exchange rate links created with trading partners as a consequence of pegging with an anchor currency. We extend their specification to include dynamics as well as separate trade creation and diversion effects for FTAs as follows

$$\log(X_{ijt}) = \beta_0 + \alpha \log(X_{ijt-1}) + \sum_{k=1}^{K} \beta_k Z_{ijt}^m + \gamma_1 CU_{ijt} + \gamma_2 DirPeg_{ijt} + \gamma_3 IndirPeg_{ijt} +$$

$$+ \gamma_4 Vol_{ijt}^{SR} + \gamma_5 Vol_{ijt}^{LR} + \sum_{m=1}^{M} \delta_m^{cr} FTA_{ijt}^m + \sum_{m=1}^{M} \delta_m^{dv} dFTA_{it}^m + \lambda_t + u_{ijt}, \tag{19}$$

where Z is a vector consisting of traditional time varying and invariant trade determinants; CU_{ijt} is binary variable that is 1 if i and j share the same currency; $DirPeg_{ijt}$ is a binary variable that is 1 if i's exchange rate is pegged to j, or vice versa

---

[16]Recent work on theoretical foundations for a dynamic gravity model specification includes Cunat and Maffezzoli (2007), Campbell (2010), and Olivero and Yotov (2011).

(but i and j are not members of the same currency union); $IndirPeg_{ijt}$ is a binary variable that takes the value of 1 if i is indirectly related to j through its peg with an anchor country; Vol_{ijt}^{SR} and Vol_{ijt}^{LR} refer to real exchange rate volatility defined over either short-run or long run horizons; FTA_{ijt} is a vector consisting of FTA dummies taking the value of 1 if i and j are members of the same FTA in a given year; FTA_{it} is a vector consisting of free trade area dummies taking the value of 1 if i is a member of an FTA in a given year; λ_t are the year-specific effects indicating common shocks across countries; and u_{ij} is the error term, $u_{ij} \sim N(0, \varsigma^2)$.

The dataset used is taken from Qureshi and Tsangarides (2010) extended to include individual entries for the free trade agreement variables from the Regional Trade Agreements database of the World Trade Organization.[17] Switching from the yearly time series to panel estimation is made possible by dividing the total period into eight-year time intervals. We potentially have a total of six panels (1960-1967, 1968-1975, 1976-1983, 1984-1991, 1992-1999, and 2000-2007), but given the inclusion of the lagged dependent variable the maximum possible number of periods is 5. Using determinants defined in (19) we identify 42 proxies and consider time effects corresponding to the span on which the data was averaged. The dataset covers 159 countries over the period 1960-2007, yielding 9,628 individual country pairs (rather than $159 \times 158/2 = 12,561$ due to missing observations). Our baseline estimation covers $9,628$ country pairs with $22,875$ observations over the period 1960-2007 with and average of 3 (out of maximum 5) observations per country pair.

C.   Results

We estimate (19) using LIBMA and compare results obtained when the same specification is estimated with (i) OLS which does not account for either model uncertainty or endogeneity; (ii) System-GMM (SGMM) which accounts for endogeneity but not model uncertainty; and (iii) Bayesian Model Averaging for linear regression models routine which accounts for model uncertainty but not endogeneity.[18] In Table 11, we present sequentially the results of estimating the dynamic gravity model using OLS, SGMM, BMA, and LIBMA. Estimated means and standard errors are presented for OLS and SGMM, and statistical significance is indicated as usual, at the 1, 5, and 10 percent levels of significance. For BMA and LIBMA, we present posterior inclusion probabilities, $p(Z_i|D)$, posterior means, and standard deviations, and identify robust those variables for which the posterior inclusion probability is above the prior (that is, $p(Z_i|D) \geq 0.50$).[19]

---

[17]In addition to the traditional gravity control variables, the dataset includes data on exchange rate regime classification and anchor currencies used to construct bilateral binary variables for currency unions, direct pegs, and indirect pegs (defined using an algorithm to associate bilateral exchange rate relations with anchor currencies). See Qureshi and Tsangarides (2010) for further details and a detailed description of the data construction and sources.

[18]The BMA estimation is performed using the Bayesian Model Averaging for linear regression models routine (bireg) by Raftery (1995) available in the open source R software http://www.r-project.org. SGMM is based on the estimator proposed by Arellano and Bond (1991), Arellano and Bover (1995) and Blundell and Bond (1998), obtained using the "xtabond2" module in Stata.

[19]Cass and Raftery (1995) further identify inclusion probability thresholds to label variables as "strongly robust," "very strongly robust," etc. suggesting stronger evidence. However, these chosen cutoffs are not strictly grounded in statistical theory and remain, therefore, merely indicative of a set of variables that we consider well estimated or robust.

Once both model uncertainty and endogeneity are properly accounted for with the use of LIBMA, results from the estimation of a dynamic trade gravity model are very different compared to those obtained using OLS, SGMM, and BMA. LIBMA identifies 25 variables out of 42 as robust, including lagged trade, several exchange rate regime variables (currency union, indirect peg, and long-run exchange rate volatility) as well as some trade creating and trade diverting FTAs. In addition, unlike any of the other three estimation methods, LIBMA also finds that the sub-Saharan dummy variable—a proxy for heterogeneity in the sample—is a robust trade determinant.

We compare each estimation method in sequence with LIBMA and interpret differences in the identified sets of determinants and estimated means as resulting from not properly accounting for model uncertainty, endogeneity, or both when OLS, SGMM, and BMA are used. First, estimation with OLS finds 33 out of the 42 variables considered to be statistically significant at least at the 10 percent level of significance (with 24 of those significant at the 1 percent level). Two potentially endogenous variables—the indirect peg and long-run volatility—are incorrectly estimated by OLS (with the former not found to be statistically significant when LIBMA labels it as robust, and the latter found to be statistically significant when LIBMA does not find it robust). In addition, the large number of the remaining variables identified as statistically significant in OLS—including several FTAs which are estimated to have both a trade creating effect as well as a trade diverting effect—disappears once model uncertainty and endogeneity are accounted for with LIBMA. Finally, estimated means are imprecisely estimated for several variables that are both identified as robust by LIBMA and found to be statistically significant by OLS, such as several of the FTAs, the currency union, and short-run volatility.

Next, we turn to SGMM. With potential endogeneity accounted for, SGMM identifies, overall, less variables as statistically significant compared to OLS. However, incorporating endogeneity alone in SGMM does not properly identify the robust determinants: comparing SGMM with LIBMA, out of the exchange rate regime and trade variables, only one (lagged trade) is found to be statistically significant in SGMM and robust in LIBMA. Similarly, several FTA variables identified as statistically significant trade determinants by SGMM receive very low inclusion probabilities once model uncertainty is properly accounted for in LIBMA. Finally, differences between BMA and LIBMA suggest that accounting for model uncertainty alone does not properly identify the trade determinants, as only about half of the LIBMA identified robust variables are also identified as robust by BMA. The application of BMA—which augments the OLS estimation by incorporating model uncertainty—identifies a subset of the OLS statistically significant variables as robust. In comparison with LIBMA, BMA fails to identify two potentially endogenous variables (indirect peg and short-run volatility) as robust, while several of the FTAs identified as robust by BMA are no longer robust when endogeneity is incorporated in LIBMA.

In summary, we find important differences in the identified bilateral trade determinants using LIBMA, compared with those identified using OLS, SGMM or BMA. We attribute these differences to the fact that LIBMA—unlike the other estimation methods presented—incorporates *both* model uncertainty and endogeneity. The results of our application underscore the importance of accounting for both endogeneity and model uncertainty in the estimation of the dynamic gravity model for trade.

VI. Conclusion

This paper proposes a limited information methodology in the context of Bayesian Model Averaging—labeled LIBMA—for panel data models where the lagged dependent variable appears as a regressor and endogenous variables appear as regressors. 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 contributes 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. The application of our LIBMA methodology to the estimation of a dynamic gravity model for trade illustrates its applicability to the case where model uncertainty and endogeneity is present in short panels. Future research could explore the possibility of using the LIBMA methodology for applications where the sample size is constrained by data availability.

References

Ahn, S.C., and P. Schmidt, 1995, Efficient estimation of models for dynamic panel data. *Journal of Econometrics* 68, 5–28.

Anderson, J. E., 1979, A theoretical foundation for the gravity equation. *American Economic Review* 69, 106–116.

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* 101, 123–64.

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

Arellano, M., and O. Bover, 1995, Another look at the instrumental-variable estimation of error components models. *Journal of Econometrics* 68, 29–52.

Back, K., and D. P. Brown, 1993, Implied probabilities in GMM estimators. *Econometrica* 61, 971–975.

Baxter, M., and M.A. Kouparitsas, 2006, What determines bilateral trade flows? *NBER Working Papers 12188*, National Bureau of Economic Research.

Bergstand, J., 1985, The gravity equation in international trade: some microeconomic foundations and empirical evidence. *Review of Economics and Statistics* 67, 474–481.

Blundell, R., and S. Bond, 1998, Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87, 114–143.

Brock, W., and S. Durlauf, 2001, Growth empirics and reality. *World Bank Economic Review* 15, 229–272.

Campbell, D., 2010, History, culture and trade: a dynamic gravity approach, Working Paper University of California at Davis.

Chernozhukov, V., and H. Hong, 2003, An MCMC approach to classical estimation. *Journal of Econometrics* 115, 293–346.

Chen, H., A. Mirestean, and C. Tsangarides, 2009, Limited Information Bayesian Model Averaging for Dynamic Panels with Short Time Periods. IMF Working Paper No. 09/74. Washington: International Monetary Fund.

Chipman, H., E.I. George, and R.E. McCulloch, 2001, The practical implementation of Bayesian model selection (with discussion). In *Model Selection* IMS Lecture Notes, Vol. 38, Lahiri P. (ed), 70–134.

Cuñat A., and M. Maffezzoli, 2007, Can comparative advantage explain the growth of US trade? *Economic Journal* 117, 583–602

Durlauf, S., and D. Quah, 1999, The new empirics of economic growth. In *Handbook of Macroeconomics* Vol. IA, Taylor J.B., Woodford M. (eds). North Holland.

Eicher, T., C. Henn, and C. Papageorgiou, 2011, Trade creation and diversion revisited: accounting for model uncertainty and natural trading partner effects. *Journal of Applied Econometrics*, forthcoming.

Eicher, T., C. Papageorgiou, and A. Raftery, 2010, Default Priors and Predictive Performance in Bayesian Model Averaging, with Application to Growth Determinants. *Journal of Applied Econometrics* 26, 30–55.

Fernàndez C., E. Ley, and M. Steel, 2001a, Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics* 16, 563–76.

Fernàndez C., E. Ley, and M.F.J. Steel, 2001b, Benchmark priors for Bayesian Model Averaging. *Journal of Econometrics* 100, 381–427.

Frankel, J., and A.K. Rose, 2002, An estimate of the effect of common currencies on trade and income. *Quarterly Journal of Economics* 117, 437–66.

Ghosh S, and S. Yamarik, 2004, Are preferential trade agreements trade creating? An application of extreme bounds analysis. *Journal of International Economics* 63, 369–395.

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* 14, 382–417.

Hong, H., and B. Preston, 2008, Bayesian averaging, prediction and nonnested model selection. *NBER Working Papers 14284*, National Bureau of Economic Research.

Jacobson T., and S. Karlsson, 2004, Finding good predictors for inflation: a Bayesian Model Averaging approach. *Journal of Forecasting* 23, 476–496.

Kass, R., and A. Raftery, 1995, Bayes factors. *Journal of the American Statistical Association* 90, 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* 90, 928–34.

Kim, J. Y., 2002, Limited information likelihood and Bayesian analysis. *Journal of Econometrics* 107, 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, 30–54.

Klein, M., and J.C. Shambaugh, 2006, Fixed exchange rates and trade, *Journal of*

International Economics 70, 359–383.

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* 73, 31–43.

Ley E., and M. Steel, 2009, On the Effect of prior assumptions in Bayesian Model Averaging with Applications to Growth Regressions. *Journal of Applied Econometrics*, 24, 651–74.

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* 89, 1335–1346.

Malik A., and J. Temple, 2009, The Geography of Output Volatility, *Journal of Development Economics*, 90, 163–178.

Mitchell, T.J., and J.J. Beauchamp, 1988, Bayesian variable selection in linear regression. *Journal of the American Statistical Association* 83, 1023–1032.

Moral-Benito, E., 2011, Determinants of economic growth: a Bayesian panel data approach. *The Review of Economics and Statistics,* forthcoming.

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, 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* 49, 169–193.

Olivero, M., and Y. Yotov, 2011, Dynamic gravity: theory and empirical implications. *Canadian Journal of Economics*, forthcoming.

Qureshi, M., and C. Tsangarides, 2010, The empirics of exchange rate regimes and trade: words vs. deeds, IMF Working Paper No. 10/48, Washington: International Monetary Fund.

Raftery, A. E., 1995, Bayesian model selection in social research. *Sociological Methodology* 25, 111–163.

Raftery, A. E., 1996, Approximate Bayes factors and accounting for model uncertainty in generalized linear models. *Biometrika* 83, 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, 179–191.

Ragusa, G., 2008, Bayesian likelihoods for moment condition models. Mimeo.

Rose, A.K., 2000, One money one market: estimating the effect of common currencies on

trade, *Economic Policy* 15, 7–46.

Sala-i-Martin, X., Doppelhofer, G., R., and R. Miller, 2004, Determinants of long-term growth: a Bayesian Averaging of Classical Estimates (BACE) approach. *American Economic Review* 94, 813–835.

Santos Silva, J.M.C., and S. Tenreyro, 2006, The Log of Gravity, *Review of Economics and Statistics* 88, 641–58.

Schennach, S.C., 2005, Bayesian exponentially tilted empirical likelihood. *Biometrika* 92, 31–46.

Temple, J., 2000, Growth Regressions and What the Textbooks Don't Tell You, *Bulletin of Economic Research*, 52, 181–205.

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, 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, 2394–2402.

Appendices

Appendix I: Representation of Moment Equations for Dynamic Panel Model

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 (6) for the FD equation

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

Similarly, W_i denotes the $(T-1) \times q(T-2)(T-1)/2$ matrix of endogenous variables to represent the instruments in (8)

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

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 to represent instruments in (9) and (10) respectively.

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

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 from moment conditions (7) and (11), respectively

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

For instruments from homoskedasticity (12) let Y_i^- be the $T \times (T-1)$ instrument matrix used for the moment conditions derived from the homoskedasticity restriction:

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

For instruments from exogenous variables (13) 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 (5).

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

Finally we define matrices U_i and G_i to summarize the moment conditions discussed so far. U_i is a $(2T-1) \times 1$ matrix defined as

$$U_i = \begin{pmatrix} u_i' & Dv_i' \end{pmatrix}'.$$

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

$$G_i = \begin{pmatrix} DX_i & DY_i & 0 & DW_i & 0 & Y_i^- \\ X_i & 0 & Y_i & 0 & W_i & 0 \end{pmatrix}.$$

Table 1: Posterior probability of the true model
LIBMA summary statistics for various N, α, and σ_v^2

α		0.95			0.50	
σ_v^2	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. $\eta_i \sim N\left(0, \sigma_\eta^2\right)$ with $\sigma_\eta^2 = 0.10$.
2. $v_{it} \sim N\left(0, \sigma_v^2\right)$.

Table 2: Posterior probability ratio of true model versus best among the rest
LIBMA summary statistics for various N, α, and σ_v^2

	α		0.95			0.50	
	σ_v^2	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. See Notes in Table 1.

Table 3: Probability of retrieving the true model
LIBMA summary statistics for various N, α, and σ_v^2

	α		0.95			0.50	
	σ_v^2	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. See Notes in Table 1.

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^2

		α = 0.95						α = 0.50					
		σ_v^2 0.05		0.10		0.20		0.05		0.10		0.20	
Sample	True Model	med	var	med	var	med	var	med	var	med	var	med	var
N=200													
yt-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
x1	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
x2	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
x3	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
x4	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
x5	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
x6	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
w1	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
w2	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													
yt-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
x1	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
x2	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
x3	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
x4	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
x5	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
x6	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
w1	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
w2	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													
yt-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
x1	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
x2	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
x3	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
x4	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
x5	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
x6	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
w1	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
w2	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: See Notes in Table 1.

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^2

| | | α 0.95 | | | | | | True | α 0.50 | | | | | |
| | | 0.05 | | 0.10 | | 0.20 | | | 0.05 | | 0.10 | | 0.20 | |
Sample	σ_v^2 True Value	med	var	med	var	med	var	Value	med	var	med	var	med	var
N=200														
yt-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
x1	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
x2	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
x3	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
x4	-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
x5	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
x6	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
w1	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
w2	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														
yt-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
x1	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
x2	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
x3	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
x4	-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
x5	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
x6	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
w1	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
w2	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														
yt-1	0.95	0.95018	0.00000	0.95017	0.00000	0.94997	0.00000	0.50	0.49532	0.00025	0.47973	0.00022	0.48687	0.00027
x1	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
x2	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
x3	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
x4	-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
x5	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
x6	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
w1	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
w2	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:See Notes in Table 1.

Table 6: Posterior probability of the true model
LIBMA summary statistics for various N, α, and σ_v^2

	α	0.95			0.50		
	σ_v^2	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.

Table 7: Posterior probability ratio of true model versus best among the rest
LIBMA summary statistics for various N, α, and σ_v^2

α		0.95			0.50		
σ_v^2	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.

Table 8: Probability of retrieving the true model
LIBMA summary statistics for various N, α, and σ_v^2

	α		0.95			0.50	
	σ_v^2	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.

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^2

		α = 0.95						α = 0.50					
		σ_v^2=0.05		σ_v^2=0.10		σ_v^2=0.20		σ_v^2=0.05		σ_v^2=0.10		σ_v^2=0.20	
Sample	True Model	med	var	med	var	med	var	med	var	med	var	med	var
N=200													
yt-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
x1	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
x2	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
x3	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
x4	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
x5	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
x6	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
w1	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
w2	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													
yt-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
x1	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
x2	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
x3	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
x4	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
x5	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
x6	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
w1	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
w2	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													
yt-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
x1	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
x2	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
x3	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
x4	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
x5	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
x6	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
w1	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
w2	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^2

| | α 0.95 | | | | | | | α 0.50 | | | | | | |
| | | σ_v^2 0.05 | | 0.10 | | 0.20 | | | 0.05 | | 0.10 | | 0.20 | |
Sample / True Value	True Value	med	var	med	var	med	var	True Value	med	var	med	var	med	var
N=200														
yt-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
x1	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
x2	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
x3	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
x4	-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
x5	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
x6	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
w1	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
w2	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														
yt-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
x1	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
x2	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
x3	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
x4	-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
x5	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
x6	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
w1	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
w2	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														
yt-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
x1	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
x2	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
x3	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
x4	-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
x5	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
x6	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
w1	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
w2	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.

Table 11. Dynamic gravity model estimation
LIBMA comparison with OLS, GMM, and BMA

	OLS			SGMM			BMA			LIBMA		
Variables	Mean	St.Error		Mean	St.Error		P(incl.)	Mean	St.Error	P(incl.)	Mean	St.Error
Regime and lagged trade												
1 Trade_{ijt-1}	0.6210	(0.0074)	***	0.3844	(0.0734)	***	1.0000	0.6252	(0.0048) +	1.0000	0.4287	(0.0192) +
2 CU_{ijt}	0.4170	(0.0808)	***	0.0595	(0.2200)		0.9960	0.3650	(0.0859) +	0.9941	0.6113	(0.2730) +
3 DirectPeg_{ijt}	0.0657	(0.0483)		0.6670	(0.5150)		0.0000	0.0000	(0.0000)	0.1637	-0.0674	(0.0319)
4 IndirectPeg_{ijt}	0.0511	(0.0489)		0.0222	(0.1760)		0.0000	0.0595	(0.0454)	0.8378	0.2916	(0.0725) +
5 VolLR_{ijt}	-0.2540	(0.0790)	***	-0.3380	(0.3310)		0.9900	-0.3844	(0.0917) +	0.0261	-0.0033	(0.0042)
6 VolSR_{ijt}	-0.1600	(0.0461)	***	-0.0848	(0.0621)		0.2630	-0.1664	(0.0606)	0.8362	-0.2828	(0.0674) +
Core gravity and heterogeneity												
7 $\log(\text{GDP}_{it}\text{GDP}_{jt})$	0.3920	(0.0101)	***	0.3785	(0.0892)	***	1.0000	0.3974	(0.0086) +	0.9963	0.7184	(0.0355) +
8 $\log(\text{GDPPC}_{it}\text{GDPPC}_{jt})$	0.0231	(0.0094)	**	0.3995	(0.1080)	***	0.1070	0.0251	(0.0088)	0.3121	0.0861	(0.0193)
9 Border_{ij}	0.2560	(0.0652)	***	0.0206	(0.2980)		0.9950	0.2776	(0.0654) +	0.8059	0.4237	(0.0898) +
10 ComColony_{ij}	0.2010	(0.0406)	***	0.7414	(0.1570)	***	1.0000	0.1972	(0.0350) +	0.9986	0.5125	(0.0648) +
11 Language_{ij}	0.1420	(0.0254)	***	-0.2440	(0.1910)		1.0000	0.1610	(0.0248) +	0.0107	0.0018	(0.0004)
12 Colony_{ij}	-0.0130	(0.3090)		0.2610	(0.5070)		0.0000	0.0000	(0.0000)	0.0502	-0.0185	(0.0375)
13 EverColony_{ij}	0.2580	(0.0515)	***	0.4470	(0.3290)		0.7520	0.2349	(0.0695) +	0.9956	0.5539	(0.0970) +
14 Island_{ij}	0.0316	(0.0230)		0.4596	(0.2590)	*	0.3090	0.0625	(0.0200)	0.8347	0.0379	(0.0357) +
15 Landlocked_{ij}	-0.1830	(0.0200)	***	0.0510	(0.1180)		1.0000	-0.1997	(0.0191) +	0.0553	-0.0063	(0.0017)
16 $\log(\text{Area}_i\text{Area}_j)$	-0.0110	(0.0049)	**	0.1695	(0.0423)	***	0.9480	-0.0206	(0.0044) +	0.9964	-0.0599	(0.0136) +
17 $\log(\text{Distance}_{ij})$	-0.5150	(0.0168)	***	-0.6323	(0.1690)	***	1.0000	-0.5128	(0.0151) +	0.9975	-0.8137	(0.0361) +
18 SSA_{ij}	-0.0324	(0.0268)		0.3500	(0.5870)		0.0700	-0.0634	(0.0229)	0.9998	0.2102	(0.0415) +
FTA Trade creation												
19 AFTA_{ijt}	0.3990	(0.1310)	***	1.2670	(2.1740)		0.0010	0.4155	(0.1912)	0.4043	0.2074	(0.1227)
20 APTA_{ijt}	0.2080	(0.2180)		4.2800	(1.4940)	***	0.0000	0.0000	(0.0000)	0.1131	0.0794	(0.0533)
21 AP_{ijt}	0.4480	(0.1840)	**	2.0460	(0.7580)	***	0.0000	0.6799	(0.3212)	0.4527	0.5452	(0.1207)
22 BILATERAL_{ijt}	-0.0138	(0.0404)		-1.1593	(0.3480)	***	0.0000	0.0000	(0.0000)	0.8461	-0.3339	(0.0932) +
23 CACM_{ijt}	0.3310	(0.1680)	**	45.5900	(36.1800)		0.0000	0.3556	(0.2439)	0.4702	0.3834	(0.0997)
24 CARICOM_{ijt}	0.1970	(0.1170)	*	1.2850	(1.0860)		0.0590	0.2901	(0.1063)	0.7702	2.8749	(0.6118) +
25 EC_{ijt}	-0.5350	(0.0737)	***	0.4650	(0.4310)		0.2440	-0.3503	(0.1491)	0.9972	-1.1764	(0.1679) +
26 EEA_{ijt}	0.2070	(0.0532)	***	-0.3150	(0.1910)	*	0.0830	0.2569	(0.0860)	0.9992	-0.4555	(0.1224) +
27 EFTA_{ijt}	-0.2200	(0.0798)	***	0.5050	(0.4290)		0.0000	-0.2632	(0.1600)	0.9996	-1.3112	(0.2805) +
28 LAIA_{ijt}	0.1680	(0.0955)	*	-0.2090	(0.9680)		0.0000	0.2545	(0.1272)	0.0331	0.0130	(0.0038)
29 MERCOSUR_{ijt}	-0.0531	(0.4110)		-17.9800	(26.7200)		0.0000	0.0000	(0.0000)	0.3755	0.7389	(0.0504)
30 NAFTA_{ijt}	-0.1660	(0.4490)		5.6510	(9.2800)		0.0000	0.0000	(0.0000)	0.3165	-0.4290	(0.1910)
FTA Trade diversion												
31 AFTA_{it}	0.4400	(0.0384)	***	1.2410	(0.8400)		1.0000	0.4455	(0.0364) +	0.9964	0.5410	(0.0734) +
32 APTA_{it}	0.3490	(0.0363)	***	2.4670	(1.4130)	*	1.0000	0.3357	(0.0383) +	0.3382	0.1545	(0.0545)
33 AP_{it}	-0.1160	(0.0665)	*	2.4030	(1.7360)		0.0000	-0.2007	(0.0552)	0.4050	0.0635	(0.0364)
34 BILATERAL_{it}	0.0461	(0.0249)	*	-0.3430	(0.2540)		0.0000	0.0549	(0.0219)	0.2297	-0.0587	(0.0116)
35 CACM_{it}	-0.1150	(0.0384)	***	1.7600	(1.4650)		0.2400	-0.1133	(0.0382)	0.7737	-0.0915	(0.0392) +
36 CARICOM_{it}	-0.1610	(0.0358)	***	-1.5530	(1.2920)		0.9990	-0.1355	(0.0307) +	0.5361	-0.0816	(0.0602) +
37 EC_{it}	0.1090	(0.0304)	***	2.4283	(0.7250)	***	0.9990	0.1176	(0.0298) +	0.5972	0.0600	(0.0388) +
38 EEA_{it}	-0.0809	(0.0335)	**	-1.3683	(0.4550)	***	0.0090	-0.0911	(0.0329)	0.9977	-0.4370	(0.0520) +
39 EFTA_{it}	-0.0857	(0.0313)	***	-1.0187	(0.5510)	*	0.0000	-0.0807	(0.0343)	0.0053	-0.0009	(0.0004)
40 LAIA_{it}	-0.1480	(0.0492)	***	-2.8664	(0.7480)	***	1.0000	-0.1901	(0.0331) +	0.9968	-0.2380	(0.0623) +
41 MERCOSUR_{it}	0.2490	(0.0615)	***	1.2450	(2.0420)		1.0000	0.3409	(0.0541) +	0.9870	0.3766	(0.0901) +
42 NAFTA_{it}	-0.1620	(0.0442)	***	-0.2750	(0.5190)		0.0000	-0.1328	(0.0590)	0.9959	-0.4475	(0.0659) +
AR(1) p-value				0.00								
AR(2) p-value				0.14								
Hansen test p-value				0.25								

Notes:

1. For OLS and SGMM asymptotic standard errors corrected for heteroskedasticity are reported in parentheses. Asterisks indicate statistical significance at 10% ("*"), 5% ("**"), and 1% ("***").

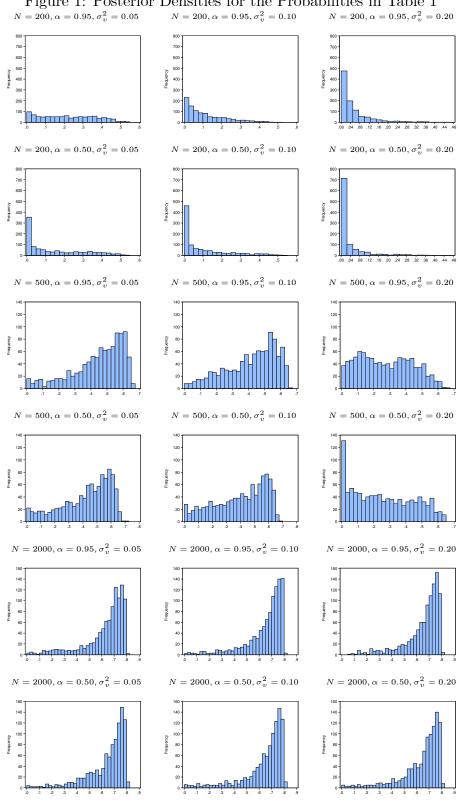2. For BMA and LIBMA "+" indicates inclusion probability above 0.50.

Figure 1: Posterior Densities for the Probabilities in Table 1

Figure 2: Posterior densities for the probabilities in Table 2

$N = 200, \alpha = 0.95, \sigma_v^2 = 0.05$ $N = 200, \alpha = 0.95, \sigma_v^2 = 0.10$ $N = 200, \alpha = 0.95, \sigma_v^2 = 0.20$

$N = 200, \alpha = 0.50, \sigma_v^2 = 0.05$ $N = 200, \alpha = 0.50, \sigma_v^2 = 0.10$ $N = 200, \alpha = 0.50, \sigma_v^2 = 0.20$

$N = 500, \alpha = 0.95, \sigma_v^2 = 0.05$ $N = 500, \alpha = 0.95, \sigma_v^2 = 0.10$ $N = 500, \alpha = 0.95, \sigma_v^2 = 0.20$

$N = 500, \alpha = 0.50, \sigma_v^2 = 0.05$ $N = 500, \alpha = 0.50, \sigma_v^2 = 0.10$ $N = 500, \alpha = 0.50, \sigma_v^2 = 0.20$

$N = 2000, \alpha = 0.95, \sigma_v^2 = 0.05$ $N = 2000, \alpha = 0.95, \sigma_v^2 = 0.10$ $N = 2000, \alpha = 0.95, \sigma_v^2 = 0.20$

$N = 2000, \alpha = 0.50, \sigma_v^2 = 0.05$ $N = 2000, \alpha = 0.50, \sigma_v^2 = 0.10$ $N = 2000, \alpha = 0.50, \sigma_v^2 = 0.20$

Notes:

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

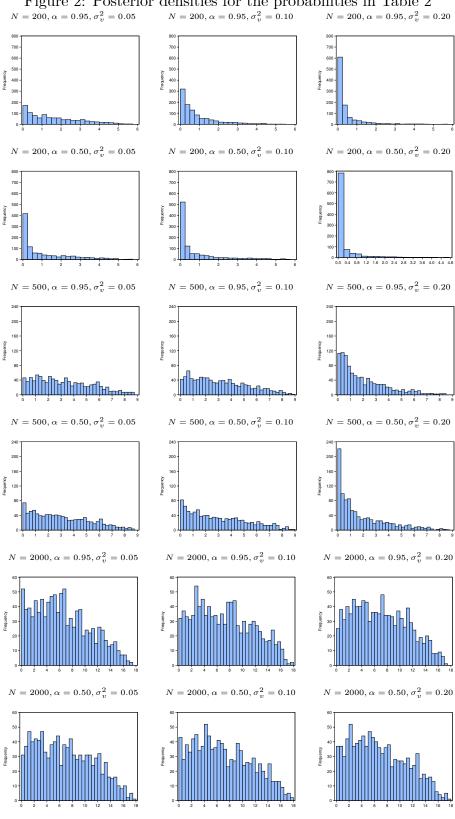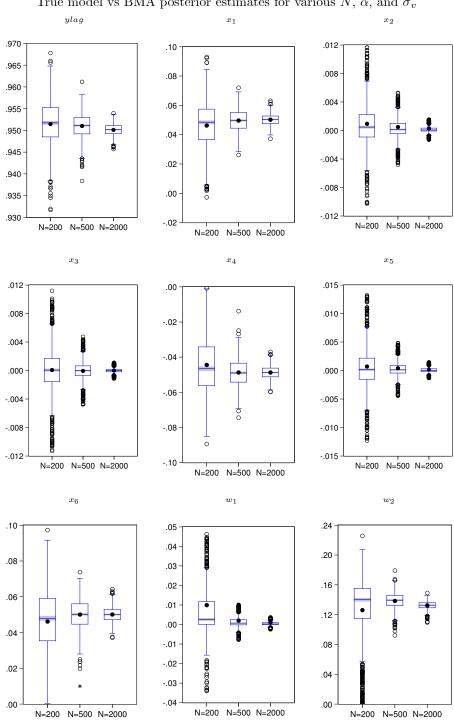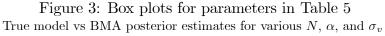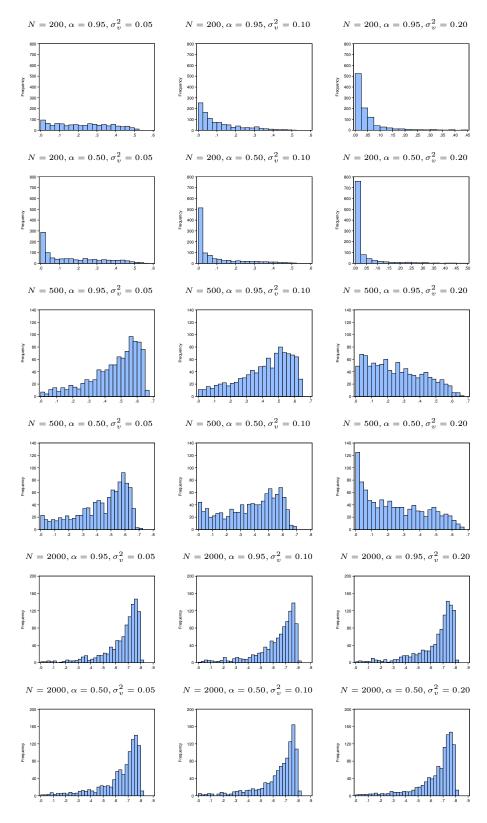2. The error term is normally distributed $v_{it} \sim N(0, \sigma_v^2)$.

Figure 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, $\eta_t \sim N(0, \sigma_\eta^2)$ where $\sigma_\eta^2 = 0.10$.

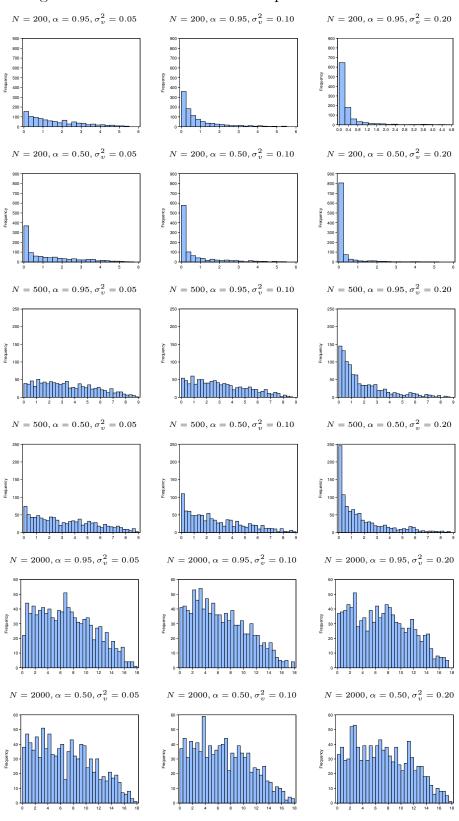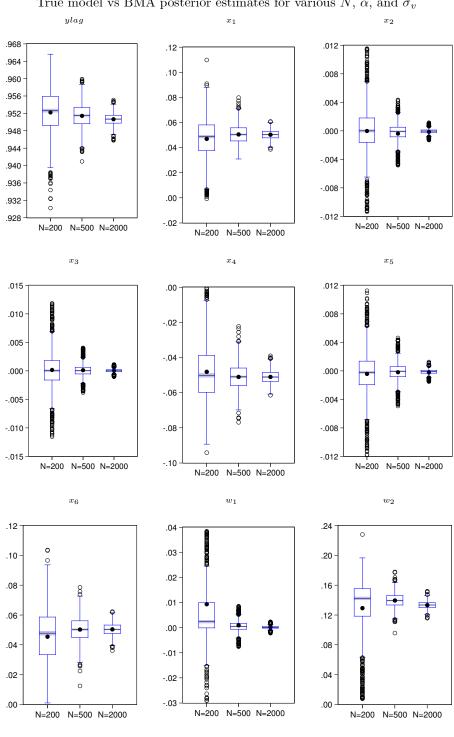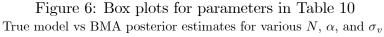2. The error term is normally distributed $v_{it} \sim N(0, \sigma_v^2)$.

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

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 IV.A.).