

Online Appendix to “Growth Determinants Revisited Using
Limited Information Bayesian Model Averaging”

Alin Mirestean	Charalambos G. Tsangarides*
International Monetary Fund	International Monetary Fund

December 2014

*Corresponding author: Research Department, International Monetary Fund, 700 19th Street, N.W., Washington D.C. 20431; Email: ctsangarides@imf.org; Tel: +1 (202) 623-5833; Fax: +1 (202) 589-5833.

Not for publication

Online-only Appendix to “Growth Determinants Revisited Using Limited Information Bayesian Model Averaging”

Appendix A. GMM estimation

Moment conditions in the GMM framework

Our set of variables includes the lagged output, a set of m exogenous variables indexed by X , a set of p predetermined variables indexed by P , and a set of q endogenous variables, indexed by W .

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 $s = 2, 3, \dots, T$. Since $y_{i,t-2}$ is not correlated with Δv_{it} it can be used as an instrument, and we have $E(y_{i,t-2}\Delta v_{it}) \neq 0$ for $t = 2, 3, \dots, T$. Moreover, since $y_{i,t-3}$ is also not correlated with Δv_{it} , $y_{i,t-3}$ can be used as an instrument if $T \geq 3$. Therefore, the following moment conditions could be used for estimation

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

Similarly, the exogenous variable x_{it}^l , $x_{it}^l \in x_{it}$ is not correlated with Δv_{it} and, therefore, it can be used as an instrument, giving additional moment conditions

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

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

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

Finally, the predetermined variable $p_{i,t-1}^l, p_{i,t-1}^l \in p_{it}$, is not correlated with Δv_{it} so it can be used as an instrument, with possible moment conditions

$$E(p_{i,t-s}^l \Delta v_{it}) = 0, \quad t = 2, 3, \dots, T; \quad s = 1, \dots, t-1;$$

$$\text{for } T \geq 2, \quad l = 1, 2, \dots, p; \quad i = 1, \dots, N.$$

From the above, 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, $q(T-2)(T-1)/2$ moment conditions for the endogenous variables, and $pT(T-1)/2$ moment conditions for the predetermined variables (Table A1).

Table A1. Moment Conditions for the First Difference Equation

Variable	Instruments	Moment conditions
$\Delta y_{i,t-1}$	$y_{i,t-2}, \dots, y_{i,0}$	$E(y_{i,t-s} \Delta v_{it}) = 0, \quad t = 2, 3, \dots, T; \quad s = 2, 3, \dots, t$
Δx_{it}^l	$x_{it}^l, \dots, x_{i1}^l$	$E(x_{it}^l \Delta v_{it}) = 0, \quad t = 2, 3, \dots, T; \quad l = 1, 2, \dots, m$
Δw_{it}^l	$w_{i,t-2}^l, \dots, w_{i,1}^l$	$E(w_{i,t-s}^l \Delta v_{it}) = 0, \quad t = 3, 4, \dots, T; \quad s = 2, 3, \dots, t-1;$ $l = 1, 2, \dots, q$
Δp_{it}	$p_{i,t-1}^l, \dots, p_{i,1}^l$	$E(p_{i,t-s}^l \Delta v_{it}) = 0, \quad t = 2, 3, \dots, T; \quad s = 1, 2, \dots, t-1;$ $l = 1, 2, \dots, p$

For the levels equation (??), first differences of the lagged dependent variable are not correlated with either the individual effects or the idiosyncratic error term. Then, we can use the following moment conditions

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

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

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

For the predetermined variables, the first difference $\Delta p_{i,t}^l$ is not correlated with u_{it} . Therefore, assuming that $p_{i,1}^l$ is observable and for $T \geq 2$, the additional moment conditions are

$$E(\Delta p_{i,t}^l u_{it}) = 0, \quad t = 2, 3, \dots, T, \quad l = 1, 2, \dots, p.$$

Finally, based on the assumptions made so far, the first difference of the exogenous variables $\Delta x_{it}^l, x_{it}^l \in x_{it}$ are not correlated with current realizations of u_{it} . Therefore, these moment conditions can be used

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

Table A2 summarizes the moment conditions for the levels equation. These 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, and $p(T - 1)$ moment conditions for the predetermined variables.

Table A2. Moment Conditions for the Levels Equation

Variable	Instruments	Moment conditions
$y_{i,t-1}$	$\Delta y_{i,t-1}$	$E(\Delta y_{i,t-1} u_{it}) = 0, \quad t = 2, 3, \dots, T$
x_{it}^l	Δx_{it}^l	$E(\Delta x_{it}^l u_{it}) = 0, \quad t = 2, 3, \dots, T; \quad l = 1, 2, \dots, m$
p_{it}^l	$\Delta p_{i,t-1}^l$	$E(\Delta p_{i,t-1}^l u_{it}) = 0, \quad t = 2, 3, \dots, T; \quad l = 1, 2, \dots, p$
w_{it}^l	$\Delta w_{i,t-1}^l$	$E(\Delta w_{i,t-1}^l u_{it}) = 0, \quad t = 3, 4, \dots, T; \quad l = 1, 2, \dots, q$

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

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

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

$\begin{pmatrix} u_{i1} & u_{i2} & \cdots & u_{iT} \end{pmatrix}'$ and $Dv_i = \begin{pmatrix} \Delta v_{i2} & \Delta v_{i3} & \cdots & \Delta v_{iT} \end{pmatrix}'$. Define a $(2T - 1) \times 1$ matrix $U_i = \begin{pmatrix} u'_i & Dv'_i \end{pmatrix}'$ that contains both the error term and the first differenced idiosyncratic random error. The full set of moment conditions can now be written in matrix form

$$E [G'_i U_i] = 0$$

where G_i is a $(2T - 1) \times \left(T + 2m - 2 + \frac{p(T+2)(T-1)}{2} + (T + 1) \left(\frac{(T-2)q+T}{2} \right) \right)$ matrix defined as

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

Reducing the number of moment conditions

The moment conditions described in G_i can be reduced by (i) limiting the lag length and/or (ii) stacking or collapsing the instrument set. We describe how this can be done for the first difference and levels equations by grouping the moment conditions for the first-difference and levels equations into matrices.

First difference equation

Recall that the first difference equation provides $T(T - 1)/2$ moment conditions for the lagged dependent variable. Simply limiting the maximum length to $\tau = 1, 2$ reduces the moment conditions to $T - 1$ and $2T - 3$, respectively. Keeping all possible lags of the dependent variable, we can reduce the count of moment conditions to $(T - 1)$ by stacking the instruments as in matrix Y_i^a

$$Y_i^a = \begin{pmatrix} y_{i0} & 0 & 0 & 0 & \cdots 0 \\ y_{i1} & y_{i0} & 0 & 0 & \cdots 0 \\ y_{i2} & y_{i1} & y_{i0} & 0 & \cdots 0 \\ y_{i3} & y_{i2} & y_{i1} & y_{i0} & \cdots 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y_{i,T_i-2} & y_{i,T_i-3} & y_{i,T_i-4} & y_{i,T_i-5} & \cdots y_{i0} \end{pmatrix}.$$

Finally, stacking and limiting the number of lags used can further reduce the count of instruments. For example, using at most 1 or 2 lags results in the $(T-1) \times 1$, and $(T-1) \times 2$ stacked matrices of instruments, Y_i^1 , and Y_i^2 , respectively

$$Y_i^1 = \begin{pmatrix} y_{i0} \\ y_{i1} \\ y_{i2} \\ y_{i3} \\ \vdots \\ y_{i,T_i-2} \end{pmatrix}, \text{ and } Y_i^2 = \begin{pmatrix} y_{i0} & 0 \\ y_{i1} & y_{i0} \\ y_{i2} & y_{i1} \\ y_{i3} & y_{i2} \\ \vdots & \vdots \\ y_{i,T_i-2} & y_{i,T_i-3} \end{pmatrix}.$$

Turning to the endogenous variables the first difference equation provides a total of $q(T-2)(T-1)/2$ moment conditions for the endogenous variables. By limiting the lag length to $\tau = 1, 2$ reduces the count of moment conditions for the endogenous variables to $q(T-2)$ and $q(2T-5)$, respectively. Alternatively, by stacking the matrix of instruments as in matrix W_i^a , the number of moment conditions is reduced to $q(T-2)$

$$W_i^a = \begin{pmatrix} 0 & 0 & 0 & \cdots 0 & \cdots 0 & 0 \\ w_{i1}^1 & 0 & 0 & \cdots w_{i1}^q & \cdots 0 & 0 \\ w_{i2}^1 & w_{i1}^1 & 0 & \cdots w_{i2}^q & \cdots 0 & 0 \\ w_{i3}^1 & w_{i2}^1 & w_{i1}^1 & \cdots w_{i3}^q & \cdots 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ w_{i,T-2}^1 & w_{i,T-3}^1 & w_{i,T-4}^1 & \cdots w_{i,T-2}^q & \cdots w_{i,2}^q & w_{i,1}^q \end{pmatrix}.$$

Further limiting the number of lags using at most 1 and 2 of all the possible lags of the endogenous variables in the stacked matrix, the matrices of instruments W_i^1 , and W_i^2 , are given by

$$W_i^1 = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ w_{i1}^1 & w_{i1}^2 & \cdots & w_{i1}^q \\ w_{i2}^1 & w_{i2}^2 & \cdots & w_{i2}^q \\ w_{i3}^1 & w_{i3}^2 & \cdots & w_{i3}^q \\ \vdots & \vdots & \vdots & \vdots \\ w_{i,T-2}^1 & w_{i,T-2}^2 & \cdots & w_{i,T-2}^q \end{pmatrix},$$

$$W_i^2 = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ w_{i1}^1 & 0 & w_{i1}^2 & \cdots & w_{i1}^q & 0 \\ w_{i2}^1 & w_{i1}^1 & w_{i2}^2 & \cdots & w_{i2}^q & w_{i1}^q \\ w_{i3}^1 & w_{i2}^1 & w_{i3}^2 & \cdots & w_{i3}^q & w_{i2}^q \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ w_{i,T-2}^1 & w_{i,T-3}^1 & w_{i,T-2}^2 & \cdots & w_{i,T-2}^q & w_{i,T-3}^q \end{pmatrix},$$

which reduces the number of moment conditions for the endogenous variables to q , and $2q$, respectively.

In a similar manner, stacking reduces the number of moment conditions for the exogenous variables from $m(T-1)$ to m . Let X_i denote the $(T-1) \times m$ matrix of instruments for the exogenous variables

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

Finally, for the predetermined variables reducing the instrument count follows the dis-

cussion on the instruments of the lagged dependent variable.¹ Simply limiting the maximum length to $\tau = 1, 2$ reduces the moment conditions to $p(T - 1)$ and $p(2T - 3)$, respectively. Stacking the instruments reduces the count of moment conditions to $p(T - 1)$, while stacking and reducing the number of lags used to $\tau = 1, 2$ limits the moment conditions to p and $2p$, respectively.

Levels equation

For the levels equation we can reduce the number of moment conditions by simply stacking the instruments. For example, we can reduce the number of moment conditions for the lagged dependent variable from $T - 1$ to 1 by just stacking the $(T - 1)$ instruments as in matrix DY_i . Similarly, for the endogenous variables, the $q(T - 2)$ moment conditions can be reduced to q by stacking the moment conditions as in the $T \times q$ instrument matrix DW_i

$$DY_i = \begin{pmatrix} 0 \\ \Delta y_{i1} \\ \Delta y_{i2} \\ \Delta y_{i3} \\ \vdots \\ \Delta y_{i,T-1} \end{pmatrix}, \quad DW_i = \begin{pmatrix} 0 & 0 \cdots & 0 \\ 0 & 0 \cdots & 0 \\ \Delta w_{i2}^1 & \Delta w_{i2}^2 \cdots & \Delta w_{i2}^q \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ \Delta w_{i,T-1}^1 & \Delta w_{i,T-1}^2 \cdots & \Delta w_{i,T-1}^q \end{pmatrix}.$$

For the exogenous variables, stacking the instruments as in the $T \times m$ matrix DX_i , reduces the number of moment conditions to m

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

¹The only difference may occur from the fact that at time $t = 0$ the predetermined variables may not have been observed and hence y_{i0} would be replaced by 0 in the instruments matrix.

Finally, let Y_i^- be the $T \times (T - 1)$ instrument matrix used for the moment conditions derived from the Ahn and Schmidt (1995) homoskedasticity restriction

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

In summary, we can reduce the $(2T - 1) \times \left(T + 2m - 2 + \frac{p(T+2)(T-1)}{2} + (T + 1) \left(\frac{(T-2)q+T}{2} \right) \right)$ G_i matrix by reducing the lags and/or stacking the instruments in the following ways:

(a) Limiting the number of instruments in G_i by including only τ maximum lags, the moment conditions in matrix form can be expressed as

$$E [G_i^{\tau'} U_i] = 0,$$

where the matrix G_i^{τ} is a $(2T - 1) \times \left((1+p) \left(\tau \left(\frac{2T-L-1}{2} \right) + T - 1 \right) + (T - 1 + 2m) + q \left(\tau \left(\frac{2T-L-3}{2} \right) + T - 2 \right) \right)$ matrix;

(b) Stacking the instruments and including all possible lags we can define the moment conditions in matrix form as

$$E [G_i^{a'} U_i] = 0,$$

where the matrix G_i^a is a $(2T - 1) \times (2m + 1 + 2p + (2 + q)(T - 1))$ matrix; and

(c) Stacking the instruments and reducing the maximum lags used to τ , results the moment conditions in matrix form

$$E [G_i^{a\tau'} U_i] = 0,$$

where the matrix $G_i^{a\tau}$ is a $(2T - 1) \times ((\tau + 1)(1 + q + p) + 2m + T - 1)$ matrix. For example, for $\tau = 1, 2$, the moment conditions in matrix form are, respectively

$$E [G_i^{a1'} U_i] = 0, \text{ or } E [G_i^{a2'} U_i] = 0,$$

where G_i^{a1} is a $(2T - 1) \times (T + 1 + 2(m + q + p))$, matrix and G_i^{a2} is a $(2T - 1) \times (T + 2 + 2m + 3(q + p))$ matrix.

To illustrate the moment reduction in practice, Table A3 below compares the number of moment conditions for values of T , m , and q under various options. In all cases presented, we assume that only one predetermined variable enters the model, the lagged dependent variable. As indicated, collapsing and/or reducing the lags yields dramatic reductions in the number of moment conditions. For example, for the case of 24 regressors (9 exogenous, one predetermined and 14 endogenous regressors) and $T = 5$, the full number of available moment conditions is 162. Limiting the number of lags used to 2 reduces the moment conditions to 145, while simply collapsing the moment conditions without limiting the number of lags reduces the number of instruments to 83. Collapsing and reducing the maximum lag length to $\tau = 2$, reduces the number of moment conditions to 67, a reduction of about 60% from the original number of 162 moment conditions. For higher values of T , the reduction in moment conditions using collapsing and/or reducing is even more dramatic.

Table A3. Moment Conditions for various options of T , m , and q

	$T = 5$			$T = 10$		
	$m = 5$	$m = 6$	$m = 9$	$m = 5$	$m = 6$	$m = 9$
Exogenous	$q = 8$	$q = 12$	$q = 14$	$q = 8$	$q = 12$	$q = 14$
Endogenous						
Uncollapsed, all lags	100	138	162	425	603	697
Uncollapsed 2 lags	89	123	145	229	323	375
Uncollapsed 1 lag	70	96	114	165	231	269
Collapsed, all lags	51	69	83	101	139	163
Collapsed 2 lags	41	55	67	46	60	72
Collapsed 1 lag	27	36	43	32	41	48

Appendix B. Derivation of the LIBIC approximation for Bayes factors

Suppose we have a strictly stationary and ergodic random process $\{\xi_i\}_{i=1}^\infty$, which takes value in the space $\Xi \subset R^s$, and a parameter space $\Theta \subset R^K$. Then, there exists a function $g : \Xi \times \Theta \rightarrow R^l$ which satisfies the following conditions

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

We further assume that the moment conditions, $E[g(\xi_i, \theta)] = 0$, hold for a unique unknown $\theta_0 \in \Theta$. Let $\hat{g}_N(\theta) = N^{-1} \sum_{i=1}^N g(\xi_i, \theta)$ denote the sample mean of the moment conditions, and assume that $S(\theta_0) \equiv E[g(\xi_i, \theta_0)g'(\xi_i, \theta_0)]$ exists and is positive definite. Furthermore $S(\theta_0) \equiv \lim_{N \rightarrow \infty} \text{Var}(N^{1/2}\hat{g}_N(\theta_0))$. Then, the following standard result holds (for a proof see Hall (2005), Lemma 3.2).

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

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

For model (??), the moment conditions for individual i can be written in the following form for model M_j

$$g(\xi_i, \theta) = g(\{y_i, x_i, w_i\}, \{\alpha, \theta_x, \theta_w\}) = G_i'(\tilde{y}_i - \tilde{z}_i C_{M_j} \theta),$$

where $\xi_i = \{\tilde{y}_i, \tilde{z}_i\}$, $\tilde{z}_i = \begin{pmatrix} \tilde{y}_{i,-1} & \tilde{x}_i & \tilde{w}_i \end{pmatrix}$, $\theta = \begin{pmatrix} \alpha & \theta_x & \theta_w \end{pmatrix}'$, while G_i is the matrix defined in (??). C_{M_j} is a $k \times k$ diagonal choice matrix with 1's on the diagonal if the corresponding variable is included in the model and 0's otherwise. We define $\theta_{M_j} = C_{M_j} \theta$

and in this notation, the parameter vector θ_{M_j} has the same dimension as θ , but with zero elements corresponding to the variables that are not included in the model M_j . The dimension of function $g(\xi_i, \theta)$ is given by the number of moment conditions considered for individual i . The vectors \tilde{y}_i and $\tilde{y}_{i,-1}$ for the dependent variable and the lagged dependent variable, respectively, are defined as follows

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

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

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

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

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

Based on the assumptions made so far, for a given model M_j the the moment conditions $E[g(\xi_i, \theta, M_j)] = 0$, hold for a unique unknown $\theta_{0, M_j} \in \Theta$. Therefore, the sample mean of

the moment conditions at θ_{0,M_j} is given by

$$\widehat{g}_N(\theta_{0,M_j}) = N^{-1} \sum_{i=1}^N G_i' \widetilde{y}_i - N^{-1} \sum_{i=1}^N G_i' \widetilde{z}_i \theta_{0,M_j}.$$

The models under consideration are panel data models where the time dimension is characterized by T , while the cross section dimension is given by N . Since $\{y_{i0}, x_i, z_i, u_i\}$ is assumed to be a strictly stationary and ergodic process with finite second moment, $E[\partial g(\xi_i, \theta_{M_j}) / \partial \theta_{M_j}] = -E[\widetilde{z}_i' G_i]$ 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, $g(\xi_i, \theta_0)$ is stationary and independent. It satisfies that $E[g(\xi_i, \theta_0)] = 0$, while $E[g(\xi_i, \theta_0)g(\xi_i, \theta_0)']$ exists and is finite positive definite. This ensures that $\lim_{N \rightarrow \infty} Var[N^{1/2} \widehat{g}_N(\theta_0)]$ exists, is finite, and is positive definite (Hansen 1982). Therefore, Lemma 1 can be applied to our dynamic panel data model.

Since we do not have any information on the form of the probability family that the model belongs to, other than the properties stated above, we assume that asymptotically the GMM estimate is the best estimate for the model. This is to say that the information on the model parameter given by the data is captured by the moment conditions asymptotically. Hence we define the likelihood $p(D|\theta, M_j)$ to be proportional to

$$p(D|\theta, M_j) \propto \exp\left(-N \frac{1}{2} \widehat{g}_N'(\theta, M_j) S^{-1}(\theta, M_j) \widehat{g}_N(\theta, M_j)\right).$$

While the moment conditions we use in the definition of the model likelihood are quite different from other non-parametric model likelihood formulations in the literature, the general functional form uses the GMM objective function and, in that sense, is similar to the construction of non-parametric likelihood functions in the existing literature (for example Kim 2002).

Then, ignoring the constants, the marginal likelihood can be expressed as

$$p(D|M_j) \propto \int_{\Theta} \exp\left(-N \frac{1}{2} \widehat{g}_N'(\theta, M_j) S^{-1}(\theta, M_j) \widehat{g}_N(\theta, M_j)\right) p(\theta|M_j) d\theta.$$

As discussed above, we assume that for a given model M_j , the associated moment conditions hold for a unique θ_{0,M_j} and we denote its estimate by $\widehat{\theta}_{0,M_j}$. In other words, $\widehat{\theta}_{0,M_j} \equiv \arg \min_{\theta} N \widehat{g}'_N(\theta, M_j) S^{-1}(\theta, M_j) \widehat{g}_N(\theta, M_j)$ and hence $\widehat{g}'_N(\theta, M_j) S^{-1}(\theta, M_j) \widehat{g}_N(\theta, M_j)$ has an absolute minimum at $\widehat{\theta}_{0,M_j}$. The prior, $p(\theta|M_j)$, is assumed to be continuous in a neighborhood of $\widehat{\theta}_{0,M_j}$ and $\exp\left(-N \frac{1}{2} \widehat{g}'_N(\theta, M_j) S^{-1}(\theta, M_j) \widehat{g}_N(\theta, M_j)\right) p(\theta|M_j)$ is integrable over the domain Θ . In addition, the first and second derivatives of $\widehat{g}'_N(\theta, M_j) S^{-1}(\theta, M_j) \widehat{g}_N(\theta, M_j)$ exist and are continuous in a neighborhood of $\widehat{\theta}_{0,M_j}$. Then using the Laplace approximation, as shown in Harding and Hausman (2007), the marginal likelihood is proportional to

$$p(D|M_j) \propto \left(\frac{2\pi}{N}\right)^{\frac{k_j}{2}} \left[\det \left(\frac{\partial^2 \left(\frac{1}{2} \widehat{g}'_N(\theta, M_j) S^{-1}(\theta, M_j) \widehat{g}_N(\theta, M_j) \right)}{\partial \theta \partial \theta'} \Big|_{\theta=\widehat{\theta}_{0,M_j}} \right) \right]^{-\frac{1}{2}} \cdot p\left(\widehat{\theta}_{0,M_j}|M_j\right) \exp\left(-N \frac{1}{2} \widehat{g}'_N\left(\widehat{\theta}_{0,M_j}\right) S^{-1}\left(\widehat{\theta}_{0,M_j}\right) \widehat{g}_N\left(\widehat{\theta}_{0,M_j}\right)\right).$$

Taking logs of both sides we have

$$\begin{aligned} \log(p(D|M_j)) &\propto -\frac{1}{2} \log \left[\det \left(\frac{\partial^2 \left(\frac{1}{2} \widehat{g}'_N(\theta, M_j) S^{-1}(\theta, M_j) \widehat{g}_N(\theta, M_j) \right)}{\partial \theta \partial \theta'} \Big|_{\theta=\widehat{\theta}_{0,M_j}} \right) \right] + \\ &\quad + \frac{k_j}{2} \log \left(\frac{2\pi}{N} \right) + \log \left(p\left(\widehat{\theta}_{0,M_j}|M_j\right) \right) \\ &\quad - N \frac{1}{2} \widehat{g}'_N\left(\widehat{\theta}_{0,M_j}\right) S^{-1}\left(\widehat{\theta}_{0,M_j}\right) \widehat{g}_N\left(\widehat{\theta}_{0,M_j}\right). \end{aligned}$$

Noting that the approximation is $O(N^{-1})$ and $\frac{\partial^2 \left(\frac{1}{2} \widehat{g}'_N(\theta, M_j) S^{-1}(\theta, M_j) \widehat{g}_N(\theta, M_j) \right)}{\partial \theta \partial \theta'} \Big|_{\theta=\widehat{\theta}_{0,M_j}}$ is a $k_j \times k_j$ matrix of order $O(1)$ due to the ergodicity assumption, and $\frac{k_j}{2} \log(2\pi)$ is also $O(1)$, the log of the marginal likelihood can be approximated by

$$\log(p(D|M_j)) \propto \frac{k_j}{2} \log \frac{1}{N} + \log \left(p\left(\widehat{\theta}_{0,M_j}|M_j\right) \right) - N \frac{1}{2} \widehat{g}'_N\left(\widehat{\theta}_{0,M_j}\right) S^{-1}\left(\widehat{\theta}_{0,M_j}\right) \widehat{g}_N\left(\widehat{\theta}_{0,M_j}\right) \quad (1)$$

We choose the unit information prior as the prior for parameters θ . The unit information prior for θ , under model M_j , is a multivariate normal distribution with mean given by the model M_j 's GMM estimate, $\widehat{\theta}_{0,M_j}$. Hence, by definition, the unit information prior

$p(\theta|M_j)$ is second order differentiable around $\hat{\theta}_{0,M_j}$ and the Laplace approximation can be used as discussed above. We also have that $\partial(p(\theta|M_j))/\partial\theta|_{\theta=\hat{\theta}_{0,M_j}} = 0$. Moreover, $p(\hat{\theta}_{0,M_j}|M_j) = (2\pi)^{-\frac{k}{2}} |\Sigma|^{-\frac{1}{2}}$, where $\Sigma = I(\hat{\theta}_{0,M_j})^{-1}$ with $I(\hat{\theta}_{0,M_j})$ being the expected Fisher information matrix evaluated at $\hat{\theta}_{0,M_j}$. Therefore $\log(p(\hat{\theta}_{0,M_j}|M_j))$ is also of order $O(1)$ and the approximation (1) becomes

$$\log(p(D|M_j)) \propto \frac{k_j}{2} \log \frac{1}{N} - N \frac{1}{2} \hat{g}'_N(\hat{\theta}_{0,M_j}) S^{-1}(\hat{\theta}_{0,M_j}) \hat{g}_N(\hat{\theta}_{0,M_j}).$$

Based on the posterior probabilities, the comparison of two models, 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). In our case the posterior odds ratio can be approximated by

$$\frac{p(M_j|D)}{p(M_i|D)} \approx \frac{p(M_j)}{p(M_i)} \exp \left(\begin{array}{c} -\frac{1}{2} N \hat{g}'_N(\hat{\theta}_{0,M_j}) S^{-1}(\hat{\theta}_{0,M_j}) \hat{g}_N(\hat{\theta}_{0,M_j}) \\ + \frac{1}{2} N \hat{g}'_N(\hat{\theta}_{0,M_i}) S^{-1}(\hat{\theta}_{0,M_i}) \hat{g}_N(\hat{\theta}_{0,M_i}) \\ + \left(\frac{k_j - k_i}{2} \log N \right) \end{array} \right), \quad (2)$$

or, in log form

$$\begin{aligned} \log \left(\frac{p(M_j|D)}{p(M_i|D)} \right) &\approx \log \frac{p(M_j)}{p(M_i)} + \left(\frac{k_j - k_i}{2} \log N \right) \\ &+ \frac{1}{2} N \hat{g}'_N(\hat{\theta}_{0,M_i}) S^{-1}(\hat{\theta}_{0,M_i}) \hat{g}_N(\hat{\theta}_{0,M_i}) \\ &- \frac{1}{2} N \hat{g}'_N(\hat{\theta}_{0,M_j}) S^{-1}(\hat{\theta}_{0,M_j}) \hat{g}_N(\hat{\theta}_{0,M_j}). \end{aligned} \quad (3)$$

For a uniform distribution over the model space we have $\frac{p(M_j|D)}{p(M_i|D)} = B_{ji}$ and (3) becomes

$$\begin{aligned} \log \left(\frac{p(M_j|D)}{p(M_i|D)} \right) &\approx \frac{1}{2} \hat{g}'_N(\hat{\theta}_{0,M_i}) S^{-1}(\hat{\theta}_{0,M_i}) \hat{g}_N(\hat{\theta}_{0,M_i}) \\ &- \frac{1}{2} N \hat{g}'_N(\hat{\theta}_{0,M_j}) S^{-1}(\hat{\theta}_{0,M_j}) \hat{g}_N(\hat{\theta}_{0,M_j}) + \left(\frac{k_j - k_i}{2} \log N \right). \end{aligned} \quad (4)$$

Recognizing that the estimate $\widehat{\theta}_{0,M_i}$ differs from model to model, the sample mean of the moment conditions for model M_j can be written as $\widehat{g}_N(\widehat{\theta}_{0,M_i}) = N^{-1} \sum_{i=1}^N G'_i(\widetilde{y}_i - \widetilde{z}_i \widehat{\theta}_{0,M_i})$. It is easy to see that G'_i , \widetilde{y}_i , and \widetilde{z}_i are the same across all models while the vector $\widehat{\theta}_{0,M_i}$ differentiates among the models by having zeros as the elements corresponding to variables not included in model M_j . In other words, the moment conditions and the observable data are the same across the universe of models,² which allows valid comparisons of posterior probabilities.

In order to approximate the Bayes factors above, we use iterative GMM estimation for $\widehat{\theta}_{0,M_i}$ with moment conditions $E \left[G'_i(\widetilde{y}_i - \widetilde{z}_i \widehat{\theta}_{0,M_i}) \right] = 0$. A consistent estimate of the weighting matrix³ is used to replace $S^{-1}(\widehat{\theta}_{0,M_i})$ in (4).

²This approach is in line with the model selection procedure proposed by Andrews and Lu (2001).

³In the case when iterated GMM is used, a consistent estimate for $S^{-1}(\widehat{\theta}_{0,M_j})$ would be $S_{(i)}^{-1}(\widehat{\theta}_{M_j,(i-1)}) = \left(N \widehat{g}_N(\widehat{\theta}_{M_j,(i-1)}) \widehat{g}'_N(\widehat{\theta}_{M_j,(i-1)}) \right)^{-1}$ where $\widehat{g}_N(\widehat{\theta}_{M_j,(i-1)})$ represents the sample moment conditions evaluated at the value of the estimate $\widehat{\theta}_{M_j,(i-1)}$ obtained in iteration $i - 1$.

Appendix C. Monte Carlo 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 $T = 5$, and vary the number of individuals, N .

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

$$\begin{pmatrix} x_{it}^1 & x_{it}^2 & x_{it}^3 & x_{it}^4 \end{pmatrix} = \begin{pmatrix} 0.3 & 0.4 & 0.8 & 0.5 \end{pmatrix} + r_t$$

with $r_t \sim N(0, I_4)$ for $t = 0, 1, \dots, T$; $i = 1, \dots, N$,

where I_4 is the four dimensional identity matrix. We allow for some correlation between the first two and the last two exogenous variables. That is, $\begin{pmatrix} x_i^5 & x_i^6 \end{pmatrix}$ 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

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

with $r_t \sim N(0, I_2)$ for $t = 0, 1, \dots, T$; $i = 1, \dots, N$,

where I_2 is the two dimensional identity matrix.

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

$$\begin{pmatrix} w_{it}^1 & w_{it}^2 \end{pmatrix} = 0.71 \begin{pmatrix} w_{i,t-1}^1 & w_{i,t-1}^2 \end{pmatrix} + 6.7v_{it} \begin{pmatrix} 1 & 1 \end{pmatrix} + r_t \text{ for } t = 1, 2, \dots, T$$

$$\begin{pmatrix} w_{i0}^1 & w_{i0}^2 \end{pmatrix} = 6.7v_{i0} \begin{pmatrix} 1 & 1 \end{pmatrix} + r_0$$

with $v_{it} \sim N(0, \sigma_v^2)$ and $r_t \sim N(0, I_2)$ for $t = 0, 1, \dots, T$.

As the data generating process for the endogenous variables indicates, the overall error term

v_{it} is assumed to be distributed normally.

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

$$y_{i0} = \frac{1}{(1-\alpha)} (x_{i0}\theta_x + w_{i0}\theta_w + \eta_i + v_{i0})$$

with $v_{i0} \sim N(0, \sigma_v^2)$ and $\eta_i \sim N(0, \sigma_\eta^2)$

where $\theta_x = \begin{pmatrix} 0.5 & 0 & 0 & -0.5 & 0 & 0.5 \end{pmatrix}'$, $\theta_w = \begin{pmatrix} 0 & 0.13 \end{pmatrix}'$, $w_{i0} = \begin{pmatrix} w_{i0}^1 & w_{i0}^2 \end{pmatrix}$, 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 $t = 1, 2, \dots, T$ the data generating process is given by

$$y_{it} = \alpha y_{i,t-1} + \theta_x x_{it} + \theta_w w_{it} + \eta_i + v_{it}$$

with $v_{it} \sim N(0, \sigma_v^2)$ and $\eta_i \sim N(0, \sigma_\eta^2)$.

Table C3b. Medians and variances of posterior inclusion probability for each variable
True model vs. LIBMA posterior inclusion probability for alpha = 0.50 and 0.30 and various N and sigma_v^2

Table with columns for Sample, True model, alpha, sigma_v^2 (0.01, 0.05, 0.10, 0.20), and sub-columns for alpha (0.50, 0.30) with Median and Variance for each. Rows are grouped by N (75, 90, 100, 150, 200, 300).

1. Two lags used stacked.

