Online Appendix A: VB for the first-stage estimations

1 The Model

The model below is model (4) in the main paper. For notational simplicity, we suppress the subscripts that indicate the model is related to the i^{th} equation of the panel SAR model.

$$Y = X\Upsilon + E, \tag{1}$$

where $vec(E) \sim N(0, \Sigma_E \otimes I_T)$. The dimensions of Y and the standardized regressors X are $T \times n$ and $T \times p$, respectively, and Υ is of dimension $p \times n$.

Let $S = E'E$ and $\Omega = \Sigma_E^{-1}$, the likelihood in equation (1) can be expressed as

$$L(\Omega) \propto |\Omega|^{\frac{T}{2}} exp\{-\frac{1}{2}tr(S\Omega)\}. \tag{2}$$

2 The Priors

2.1 The Priors of Υ

Let $\gamma = vec(\Upsilon)$. Following Bhattacharya, Pati, Pillai, and Dunson (2015), we set hierarchical D-L prior for the j^{th} ($j = 1, ..., np$) element of γ as follows:

$$\gamma_j|\phi, \tau \sim DE(\phi_j\tau), \;\; \phi_j \sim Dir(a, ..., a). \tag{3}$$

Conditional prior for γ_j is $DE(\phi_j\tau)$ implies a zero mean Double Exponential or Lapalace distribution with the density $f(\gamma_j) = (2\phi_j\tau)^{-1}exp(-\frac{|\gamma_j|}{\phi_j\tau})$ for $\gamma_j \in \mathbb{R}$.

Next, we set a Gamma prior for τ:

$$\tau \sim G(npa, 1/2). \tag{4}$$

The above hierarchical D-L prior for γ_j can be expressed as:

$$\gamma_j \sim N(0, \psi_j\phi_j^2\tau^2), \;\; \psi_j \sim Exp(1/2). \tag{5}$$

Hence, the prior of γ is $N(\mathbf{0}, \underline{V})$, where $\underline{V} = diag(\psi_1\phi_1^2\tau^2, ..., \psi_{np}\phi_{np}^2\tau^2)$.

2.2 The Priors of Ω

We set Exponential priors and D-L priors for the elements of Ω as follows:

$$
\begin{aligned}
&\omega_{ii} \sim Exp(\underline{s}), \;\; i = 1, ..., n \\
&\omega_{ij} \sim N(0, \psi_{\omega,ij}\phi^2_{\omega,ij}\tau^2_{\omega}), \;\; \psi_{\omega,ij} \sim Exp(1/2), \;\; i < j = 2, ..., n \\
&\phi_{\omega,ij} \sim Dir(a_{\omega}, ..., a_{\omega}), \;\; \tau_{\omega} \sim G(\frac{n^2 - n}{2} a_{\omega}, 1/2)
\end{aligned}
\tag{6}
$$

where with a slight abuse of notations, we use ω_{ii} and ω_{ij} to denote the diagonal and off-diagonal elements of Ω.

3 The Posteriors

3.1 The Posteriors of B and its related hyperparameters

The conditional posterior of γ is $N(\tilde{\gamma}, \tilde{V})$, where $\tilde{V} = (\underline{V}^{-1} + \Omega \otimes (X'X))^{-1}$, and $\tilde{\gamma} = \tilde{V}(\Omega \otimes X')vec(Y)$.

The conditional posterior of τ is $giG(npa - np, 1, \sum_{j=1}^{np} \frac{2|\gamma_j|}{\phi_j})$. [1]

The conditional posterior of ϕ_j can be derived as following: First, we have $\xi_j \sim giG(a - 1, 1, 2|\gamma_j|)$. Next let et $\Xi = \sum_{j=1}^{np} \xi_j$. The conditional posterior of ϕ_j can then be found to be ξ_j/Ξ.

The conditional posterior of $1/\psi_j$ is Inverse Gaussian with mean $\sqrt{\frac{\phi_j^2\tau^2}{\gamma_j^2}}$ and scale parameter 1.

3.2 The Posteriors of Ω and its related hyperparameters

Following Wang (2012), we focus on the last column and row of Ω then use Block Gibbs sampler to update the relevant parameters and hyperparameters.

Let H be the $n \times n$ matrix with 0 diagonal elements and the off diagonal element at i^{th} row and j^{th} column be $\psi_{\omega,ij}\phi^2_{\omega,ij}\tau^2_{\omega}$. Partition Ω, S and H as follows:

$$
\Omega = \begin{pmatrix} \Omega_{-n,-n} & \omega_{-n,n} \\ \omega'_{-n,n} & \omega_{nn} \end{pmatrix} \;\; S = \begin{pmatrix} S_{-n,-n} & s_{-n,n} \\ s'_{-n,n} & s_{nn} \end{pmatrix} \;\; H = \begin{pmatrix} H_{-n,-n} & h_{-n,n} \\ h'_{-n,n} & h_{nn} \end{pmatrix} \tag{7}
$$

[1] $y \sim giG(p, a, b)$ if $f(y) \propto y^{p-1} exp[-\frac{1}{2}(ay + b/y)]$.

where $-n$ denotes the set of all indices except for n.

Let $b_1 = \omega_{n,n} - \omega'_{-n,n}\Omega^{-1}_{-n,-n}\omega_{-n,n}$ and $b_2 = \omega_{-n,n}$. The conditional posteriors are as follows:

$$\begin{aligned}
&b_1 \sim G(\frac{T}{2}+1, \frac{s_{nn}+\underline{s}}{2}) \\
&b_2 \sim N(-C\mathbf{s}_{-n,n}, C), \;\; C = ((\underline{s}+s_{nn})\Omega^{-1}_{-n,-n} + H^{*-1})^{-1}, \;\; H^* = diag(h_{-n,n}) \\
&1/\psi_{\omega,ij} \sim iG(\sqrt{\frac{\tau^2_\omega \phi^2_{\omega,ij}}{\omega^2_{ij}}}, 1) \\
&\tau_\omega \sim giG(\frac{n^2-n}{2}(a_3-1), 1, \sum_{i<j}\frac{2|\omega_{ij}|}{\phi_{\omega,ij}}) \\
&\xi_{\omega,ij} \sim giG(a_\omega - 1, 1, 2|\omega_{ij}|), \;\; \phi_{\omega,ij} = \xi_{\omega,ij}/\sum_{i<j}\xi_{\omega,ij}
\end{aligned} \tag{8}$$

4 Optimal VB q densities

4.1 $q(\gamma)$

$$q(\gamma) \sim N(\bar{\gamma}, \overline{V}), \tag{9}$$

where

$$\overline{V} = (V^{-1} + \overline{\Omega} \otimes (X'X))^{-1},$$

$$\overline{\gamma} = \overline{V}(\overline{\Omega} \otimes X')vec(Y),$$

and

$$V^{-1} = diag(\frac{1}{\overline{\psi_1}\,\overline{\phi^2_1}\,\overline{\tau^2}}, ..., \frac{1}{\overline{\psi_{np}}\,\overline{\phi^2_{np}}\,\overline{\tau^2}}).$$

4.2 $q(\tau)$

$$q(\tau) \sim giG[npa - np, 1, \sum_{j=1}^{np} 2(\overline{\gamma}^2_j + \overline{V}_{jj})^{1/2}\frac{1}{\overline{\phi_j}}] \tag{10}$$

Let $\chi = \sum_{j=1}^{np} 2(\overline{b}^2_j + \overline{V}_{jj})^{1/2}\frac{1}{\overline{\phi_j}}$. We have

$$\bar{\tau} = \frac{\sqrt{\chi}K_{npa-np+1}(\sqrt{\chi})}{K_{npa-np}(\sqrt{\chi})},$$

and

$$\overline{\tau^2} = \bar{\tau}^2 + \chi[\frac{K_{npa-np+2}(\sqrt{\chi})}{K_{npa-np}(\sqrt{\chi})} - (\frac{K_{npa-np+1}(\sqrt{\chi})}{K_{npa-np}(\sqrt{\chi})})^2],$$

where $K_*[\bullet]$ is the modified Bessel functions of the second kind.

4.3 $q(\psi_j)$

$$q(\psi_j^{-1}) \sim iG(\sqrt{\frac{\overline{\phi_j^2}\,\overline{\tau^2}}{\overline{\gamma}_j^2 + \overline{V}_{jj}}}, 1), \tag{11}$$

Let $\rho = \sqrt{\frac{\overline{\phi_j^2}\,\overline{\tau^2}}{\overline{\gamma}_j^2 + \overline{V}_{jj}}}$.

$$\overline{\psi_j^{-1}} = \rho,$$

and

$$\overline{\psi}_j = 1 + 1/\rho.$$

4.4 $q(\phi_j)$

$$q(\xi_j) \sim giG(a-1, 1, 2\sqrt{\overline{\gamma}_j^2 + \overline{V}_{jj}}) \tag{12}$$

Let $\varpi = 2\sqrt{\overline{\gamma}_j^2 + \overline{V}_{jj}}$. We have

$$\overline{\xi_j} = \frac{\sqrt{\varpi} K_a(\sqrt{\varpi})}{K_{a-1}(\sqrt{\varpi})},$$

and

$$var(\xi_j) = \varpi\{\frac{K_{a+1}(\sqrt{\varpi})}{K_{a-1}(\sqrt{\varpi})} - [\frac{K_a(\sqrt{\varpi})}{K_{a-1}(\sqrt{\varpi})}]^2\},$$

where $var(\bullet)$ denotes the variance.

Scaling ξ_i, we have

$$\overline{\phi_j} = \frac{\overline{\xi_j}}{\sum_{j=1}^{np} \overline{\xi_j}},$$

and

$$\overline{\phi_j^2} = \overline{\phi_j}^2 + \frac{var(\xi_j)}{(\sum_{j=1}^{np} \overline{\xi_j})^2}.$$

Thus, the optimal q density of $\phi_{i,j}$ takes the following form:

$$q(\phi_j) \sim giG(a-1, \sum_{j=1}^{np} \overline{\xi_j}, \frac{2\sqrt{\overline{\gamma}_j^2 + (\overline{V}_{jj})^2}}{\sum_{j=1}^{np} \overline{\xi_j}}) \tag{13}$$

4.5 $q(b_1)$

$$q(b_1) \sim G(\frac{T}{2}, \overline{s}_{n,n}), \tag{14}$$

where

$$\overline{s}_{n,n} = \frac{1}{2}(s_{n,n} + tr(X'X\overline{V_n}) + \underline{s}),$$

and

$$\overline{V_n} = V_{(n-1)\times p+1:n\times p,(n-1)\times p+1:n\times p}.$$

Hence

$$\overline{b_1} = \frac{\frac{T}{2}}{\overline{s}_{n,n}}$$

4.6 $q(b_2)$

Let $\overline{s}_{-n,n} = s_{-n,n} + \tilde{s}_{-n,n}$, where $\tilde{s}_{-n,n}$ is a $(n-1)\times 1$ vector with the j^{th} element being $tr(X'XA_j)$ and $A_j = V_{(j-1)\times p+1:j\times p,(j-1)\times p+1:j\times p}$. [2]

$$q(b_2) \sim N(-\overline{C}\overline{s}_{-n,n}, \overline{C}), \tag{15}$$

where

$$\overline{C} = (2\overline{s}_{n,n}\Omega^{-1}_{-n.-n} + \overline{H}^{*-1})^{-1},$$

and

$$\overline{b_2} = (-\overline{C}\overline{s}_{-n,n}).$$

Note that $\overline{H}^* = diag(\overline{h}_{-n,n})$, and j^{th} element of $\overline{h}_{-n,n}$ is $\overline{\psi_{\omega,jn}}\overline{\phi^2_{\omega,jn}}\overline{\tau^2_\omega}$.

4.7 $q(\tau_\omega)$

$$q(\tau_\omega) \sim giG[\frac{n^2-n}{2}(a_\omega - 1), 1, \sum_{j<k} 2(\overline{\omega_{jk}}^2 + \overline{C}_{jj})^{1/2}\frac{1}{\overline{\overline{\phi_{\omega,jk}}}}],^3 \tag{16}$$

[2] To calculate $\overline{s}_{-i,i}$, for $i \neq n$, we need to delete the relevant $(i-1)\times p+1$ to $i\times p$ rows and $(i-1)\times p+1$ to $i\times p$ columns of $\overline{V}$ to construct A_j.

[3] Note that $\overline{C}_{jj}$ changes when k changes.

Let $\chi_\omega = \sum_{j<k} 2(\overline{\omega_{jk}}^2 + \overline{V}_{jj})^{1/2}(\overline{\phi_{\omega,jk}})^{-1}$. We have

$$\overline{\tau_\omega} = \frac{\sqrt{\chi_\omega} K_{\frac{n^2-n}{2}(a_\omega-1)+1}(\sqrt{\chi_\omega})}{K_{\frac{n^2-n}{2}(a_\omega-1)}(\sqrt{\chi_\omega})},$$

and

$$\overline{\tau_\omega^2} = \overline{\tau_\omega}^2 + \chi_\omega[\frac{K_{\frac{n^2-n}{2}(a_\omega-1)+2}(\sqrt{\chi})}{K_{\frac{n^2-n}{2}(a_\omega-1)}(\sqrt{\chi_\omega})} - (\frac{K_{\frac{n^2-n}{2}(a_\omega-1)+1}(\sqrt{\chi_\omega})}{K_{\frac{n^2-n}{2}(a_\omega-1)}(\sqrt{\chi_\omega})})^2].$$

4.8 $q(\psi_{\omega,jn})$

$$q(\psi_{\omega,jn}^{-1}) \sim iG(\sqrt{\frac{\overline{\phi_{\omega,jn}^2}\,\overline{\tau_\omega^2}}{\overline{\omega_{jn}^2} + \overline{C}_{jj}}}, 1) \tag{17}$$

Let $\rho_\omega = \sqrt{\frac{\overline{\phi_{\omega,jn}^2}\,\overline{\tau_\omega^2}}{\overline{\omega_{jn}}^2 + \overline{C}_{jj}}}$.

$$\overline{\psi_{\omega,jn}^{-1}} = \rho_\omega,$$

and

$$\overline{\psi_{w,jn}} = 1 + 1/\rho_\omega.$$

4.9 $q(\phi_{\omega,jn})$

$$q(\xi_{\omega,jn}) \sim giG(a_\omega - 1, 1, 2\sqrt{\overline{\omega_{jn}}^2 + \overline{C}_{jj}}) \tag{18}$$

Let $\varpi_\omega = 2\sqrt{\overline{\omega_{jn}}^2 + \overline{C}_{jj}}$. We have

$$\overline{\xi_{\omega,jn}} = \frac{\sqrt{\varpi_\omega} K_{a_\omega}(\sqrt{\varpi_\omega})}{K_{a_\omega-1}(\sqrt{\varpi_\omega})},$$

and

$$var(\xi_{\omega,jn}) = \varpi_\omega\{\frac{K_{a_\omega+1}(\sqrt{\varpi_\omega})}{K_{a_\omega-1}(\sqrt{\varpi_\omega})} - [\frac{K_{a_\omega}(\sqrt{\varpi_\omega})}{K_{a_\omega-1}(\sqrt{\varpi_\omega})}]^2\},$$

where $var(\bullet)$ denotes the variance.

Scaling $\xi_{\omega,jn}$, we have

$$\overline{\phi_{\omega,jn}} = \frac{\overline{\xi_{\omega,jn}}}{\sum_{j<k} \overline{\xi_{\omega,jk}}},$$

and

$$\overline{\phi^2_{\omega,jn}} = \overline{\phi_{\omega,jn}}^2 + \frac{var(\xi_{\omega,jn})}{(\sum\limits_{j<k} \overline{\xi_{\omega,jk}})^2}.$$

Thus, the optimal q density of $\phi_{\omega,ij}$ takes the following form:

$$q(\phi_{\omega,jn}) \sim giG(a-1, \sum_{j<k} \overline{\xi_{\omega,jk}}, \frac{2\sqrt{\overline{\omega_{jn}}^2 + (\overline{C}_{jj})^2}}{\sum\limits_{j<k} \overline{\xi_{\omega,jk}}}) \tag{19}$$

5 ELBO

The evidence lower bound ($ELBO$) is as follows:[4]

$$\begin{aligned} ELBO &= E\{\log p(Y, \gamma, \tau, \psi, \phi, b_1, b_2, \tau_\omega, \psi_\omega, \phi_\omega)\} - E\{\log q(\gamma, \tau, \psi, \phi, b_1, b_2 \tau_\omega, \psi_\omega, \phi_\omega)\} \\ &= E\{\log p(Y|\gamma, \tau, \psi, \phi, b_1, b_2 \tau_\omega, \psi_\omega, \phi_\omega)\} + E\{\log p(\gamma)\} + \sum E\{\log p(b_1)\} + \sum E\{\log p(b_2)\} \\ &+ E\{\log p(\phi)\} + E\{\log p(\psi)\} + E\{\log p(\tau)\} + E\{\log p(\phi_\omega)\} + E\{\log p(\psi_\omega)\} + E\{\log p(\tau_\omega)\} \\ &- E\{\log q(\gamma)\} - \sum E\{\log q(b_1)\} - \sum E\{\log q(b_2)\} - E\{\log q(\phi)\} - E\{\log q(\psi)\} - E\{\log q(\tau)\} \\ &- E\{\log q(\phi_\omega)\} - E\{\log q(\psi_\omega)\} - E\{\log q(\tau_\omega)\}, \end{aligned} \tag{20}$$

where

$$\begin{aligned} &E\{\log p(Y|\gamma, \tau, \psi, \phi, b_1, b_2, \tau_\omega, \psi_\omega, \phi_\omega)\} = \\ &-\frac{T}{2} log|\overline{\Sigma}| - \frac{1}{2}\left[vec(\overline{\Omega})\right]' vec(Y'Y - 2XBY' + XBB'X' + \sum_{j=1}^{n} XK_jX') + Const., \end{aligned} \tag{21}$$

with the elements of K_j retrieved from appropriate rows and columns of $\overline{V}$.

$$E\{\log p(\gamma)\} = -\frac{1}{2}E[log(|V|)] - \frac{1}{2}\left[\overline{\gamma}'V^{-1}\overline{\gamma} + tr(V^{-1}\overline{V})\right] + Const., \tag{22}$$

where $E[log(|V|)] = \sum_{j=1}^{np} E[\log(\psi_j) + 2\log(\phi_j)] + 2npE[\log(\tau)]$, $E[\log(\psi_j) + 2\log(\phi_j)] = \int_0^\infty q(\psi_j)\log(\psi_j)d\psi_j + 2\int_0^\infty q(\phi_j)\log(\phi_j)d\phi_j$ and $E[\log(\tau)] = \int_0^\infty q(\tau)\log(\tau)d\tau$.

[4]It is seen that the $ELBO$ described below contains many constant terms. Those terms need to be dropped from $ELBO$s in VB iterations to speed up the algorithm.

$$E\{\log p(b_1)\} = -(\underline{\nu} - 1)log(\overline{s}_{n,n}) - \underline{s}\overline{b_1} + Const., \tag{23}$$

$$E\{\log p(b_2)\} = -\frac{1}{2}E(log(|H^*|)) - \frac{1}{2}\left[\overline{b_2}'H^*\overline{b_2} + tr('H^{*-1}\overline{C})\right] + Const., \tag{24}$$

where $E[log(|H^*|)] = \sum_{j=1}^{n-1} E[\log(\psi_{\omega,ij}) + 2\log(\phi_{\omega,ij})] + 2(n-1)E[\log(\tau_\omega)]$, $E[\log(\psi_{\omega,ij}) + 2\log(\phi_{\omega,ij})] = \int_0^\infty q(\psi_{\omega,ij})\log(\psi_{\omega,ij})d\psi_{\omega,ij} + 2\int_0^\infty q(\phi_{\omega,ij})\log(\phi_{\omega,ij})d\phi_{\omega,ij}$ and $E[\log(\tau_\omega)] = \int_0^\infty q(\tau_\omega)\log(\tau_\omega)d\tau_\omega$.

$$E\{\log p(\tau)\} = -npa\left[\int_0^\infty (q(\tau)\log\tau)\,d\tau\right] - 0.5\overline{\tau} + Const., \tag{25}$$

$$E\{\log p(\psi)\} = -0.5\sum_{j=1}^{np}\overline{\psi_j} + Const., \tag{26}$$

$$E\{\log p(\phi)\} = -(a-1)\sum_{j=1}^{np}\left[\int_0^\infty (q(\phi_j)\log\phi_j)d\phi_j\right], \tag{27}$$

$$E\{\log p(\tau_\omega)\} = -\frac{n^2-n}{2}a_\omega\left[\int_0^\infty (q(\tau_\omega)\log\tau_\omega)\,d\tau_\omega\right] - 0.5\left[\int_0^\infty (q(\tau_\omega)\tau_\omega)\,d\tau_\omega\right] + Const., \tag{28}$$

$$E\{\log p(\psi_\omega)\} = -0.5\sum_{j=1}^{(n^2-n)/2}\int_0^\infty (\psi_{\omega,j}q(\psi_{\omega,j}))d\psi_{\omega,j} + Const., \tag{29}$$

$$E\{\log p(\phi_\omega)\} = -(a_\omega - 1)\sum_{j=1}^{(n^2-n)/2}\left[\int_0^\infty (q(\phi_{\omega,j})\log\phi_{\omega,j})d\phi_{\omega,j}\right], \tag{30}$$

$$E\{\log q(\gamma)\} = -\frac{1}{2}log(|\overline{V}|) + Const., \tag{31}$$

$$E\{\log p(b_1)\} = log(\overline{s}_{n,n}) + Const., \tag{32}$$

$$E\{\log p(b_2)\} = -\frac{1}{2}E(log(\overline{C})) + Const., \tag{33}$$

$$E\{\log q(\tau)\} = \int_0^\infty q(\tau)\log q(\tau) d\tau, \tag{34}$$

$$E\{\log q(\psi)\} = \sum_{j=1}^{np} \int_0^\infty q(\psi_j)\log q(\psi_j) d\psi_j, \tag{35}$$

$$E\{\log q(\phi)\} = \sum_{j=1}^{np} \int_0^\infty q(\phi_j)\log q(\phi_j) d\phi_j, \tag{36}$$

$$E\{\log q(\tau_\omega)\} = \int_0^\infty q(\tau_\omega)\log q(\tau_\omega) d\tau_\omega, \tag{37}$$

$$E\{\log q(\psi_\omega)\} = \sum_{j=1}^{(n^2-2)/2} \int_0^\infty q(\psi_{\omega,j})\log q(\psi_{\omega,j}) d\psi_{\omega,j}, \tag{38}$$

and

$$E\{\log q(\phi_\omega)\} = \sum_{j=1}^{(n^2-2)/2} \int_0^\infty q(\phi_{\omega,j})\log q(\phi_{\omega,j}) d\phi_{\omega,j}. \tag{39}$$

Terms in ELBO that do not have clear analytical forms need to be numerically estimated.