Online Appendix To: Bayesian Doubly Adaptive Elastic-Net Lasso For VAR Shrinkage

Deborah Gefang*
Department of Economics
University of Lancaster
email: d.gefang@lancaster.ac.uk

April 17, 2013

*I would like to thank Gary Koop, Esther Ruiz and two anonymous referees for their constructive comments. I would also like to thank the conference participants of CFE11, ESEM2012, and RCEF2012 for helpful discussions. Any remaining errors are my own responsibility.

1 Technical Details for Models Nested in DAE-Lasso

This section presents the priors, posteriors, and full conditional Gibbs schemes for Lasso, adaptive Lasso, e-net Lasso, and adaptive e-net Lasso.

1.1 Lasso VAR Shrinkage

Following Song and Bickel (2011), we define Lasso estimator for a VAR as:

$$\hat{\beta}_L = arg\ min_\beta \{[y-(I_n \otimes X)\beta]'[y-(I_n \otimes X)\beta] + \lambda_1 \sum_{j=1}^{N^2k} |\beta_j|\} \tag{1}$$

Correspondingly, the conditional multivariate mixture prior for β takes the following form:

$$\begin{aligned} \pi(\beta|\Sigma,\Gamma,\lambda_1) \propto \prod_{j=1}^{N^2k} \{ \int_0^\infty \frac{1}{\sqrt{2\pi f_j(\Gamma)}} exp[-\frac{1}{2f_j(\Gamma)}\beta_j^2] d(f_j(\Gamma))\} \\ \times \{|M|^{-\frac{1}{2}} exp(-\frac{1}{2}\Gamma' M^{-1}\Gamma)\}^2 \end{aligned} \tag{2}$$

where $\Gamma = [\gamma_1, \gamma_2, ..., \gamma_{N^2k}]'$, $M = \Sigma \otimes I_{Nk}$, and $f_j(\Gamma)$ is a function of Γ and Λ_1 to be defined later. In this mixture prior, the terms associated with the L_1 penalty are conditional on Σ through $f_j(\Gamma)$. In equation (2), the variances of β_a and β_b for $a \neq b$ are related through M. However, β_a and β_b themselves are independent of each other.

We need to find an appropriate $f_j(\Gamma)$ which provides us tractable posteriors. The last term in equation (2) takes the form of a multivariate Normal distribution $\Gamma \sim N(0, M)$. For ease of exposition, we first write the

$N^2k \times N^2k$ covariance matrix M as following:

$$M = \begin{pmatrix} M_{1,1} & ... & M_{1,j} & M_{1,j+1} & ... & M_{1,N^2k} \\ ... & ... & ... & ... & ... & ... \\ M_{j,1} & ... & M_{j,j} & M_{j,j+1} & ... & M_{j,N^2k} \\ M_{j+1,1} & ... & M_{j+1,j} & M_{j+1,j+1} & ... & M_{j+1,N^2k} \\ ... & ... & ... & ... & ... & ... \\ M_{N^2k,1} & ... & M_{N^2k,j} & M_{N^2k,j+1} & ... & M_{N^2k,N^2k} \end{pmatrix} \tag{3}$$

Let $H_j = (M_{j,j+1}, ..., M_{j,N^2k}) \begin{pmatrix} M_{j+1,j+1} & ... & M_{j+1,N^2k} \\ ... & ... & ... \\ M_{N^2k,j+1} & ... & M_{N^2k,N^2k} \end{pmatrix}^{-1}$.

We next construct independent variables τ_j for $j = 1, 2, ..., N^2k$ using standard textbook techniques (e.g. Anderson, 2003; Muirhead 1982).

$$\tau_1 = \gamma_1 + H_1(\gamma_2, \gamma_3, ..., \gamma_{N^2k})' \tag{4}$$

$$\tau_2 = \gamma_2 + H_2(\gamma_3, \gamma_4, ..., \gamma_{N^2k})' \tag{5}$$

$$...$$

$$\tau_{N^2K-1} = \gamma_{N^2k-1} + H_{N^2k-1}\gamma_{N^2k} \tag{6}$$

$$\tau_{N^2K} = \gamma_{N^2k} \tag{7}$$

The joint density of $\tau_1, \tau_2, ..., \tau_{N^2k}$ is

$$N(\tau_1|0, \sigma^2_{\gamma_1})N(\tau_2|0, \sigma^2_{\gamma_2})...N(\tau_{N^2k}|0, \sigma^2_{\gamma_{N^2k}}) \tag{8}$$

where $\sigma^2_{\gamma_j} = M_{j,j} - H_j(M_{j,j+1}, ..., M_{j,N^2k})'$, with $\sigma^2_{\gamma_{N^2k}} = M_{N^2k,N^2k}$. Note that it is computationally feasible to derive $\sigma^2_{\gamma_j}$ when M is sparse.

The Jacobian of transforming $\Gamma \sim N(0, M)$ to (8) is 1. Defining $\eta_j = \tau_j/\lambda_1$, we can write (8) as

$$N(\eta_1|0, \sigma^2_{\gamma_1}\lambda_1^{-2})N(\eta_2|0, \sigma^2_{\gamma_2}\lambda_1^{-2})...N(\eta_{N^2k}|0, \sigma^2_{\gamma_{N^2k}}\lambda_1^{-2}) \tag{9}$$

Let $f_j(\Gamma) = 2(\eta_j^2)$, the scale mixture prior is:

$$\begin{aligned}\pi(\beta|\Sigma, \Gamma, \lambda_1) \propto \prod_{j=1}^{N^2k} \Big\{ \int_0^\infty \frac{1}{\sqrt{2\pi(2\eta_j^2))}} exp[-\frac{\beta_j^2}{2(2\eta_j^2)}]d(2\eta_j^2) \\ \times \frac{\lambda_1^2}{2\sigma^2_{\gamma_j}} exp[-\frac{1}{2}\frac{2\eta_j^2}{(\sigma^2_{\gamma_j})/\lambda_1^2}] \Big\}\end{aligned} \tag{10}$$

The last two terms in (10) constitute a scale mixture of Normals (with an exponential mixing density), which can be expressed as the univariate Laplace distribution $\frac{\lambda_1}{2\sqrt{\sigma^2_{\gamma_j}}} exp(-\frac{\lambda_1}{\sqrt{\sigma^2_{\gamma_j}}}|\beta_j|)$.

Equation (10) shows that the conditional prior for β_j is $N(0, \frac{1}{2\eta_j^2})$, and the conditional prior for β is

$$\beta|\Gamma, \Sigma, \Lambda_1, \Lambda_2 \sim N(0, D^*_\Gamma) \tag{11}$$

where $D^*_\Gamma = diag([\frac{1}{2\eta_1^2}, \frac{1}{2\eta_2^2}, ..., \frac{1}{2\eta^2_{N^2k}}])$.

Priors for Σ and λ_1^2 can be elicited following standard practice in VAR and Lasso literature. In this paper, we set Wishart prior for Σ^{-1} and Gamma prior for λ_1^2: $\Sigma^{-1} \sim W(\underline{S}^{-1}, \underline{\nu})$, $\lambda_1^2 \sim G(\underline{\mu}_{\lambda_1^2}, \underline{\nu}_{\lambda_1^2})$.

The full conditional posterior for β is $\beta \sim N(\overline{\beta}, \overline{V}_\beta)$, where $\overline{V}_\beta = [(I_N \otimes X)^{'}(\Sigma^{-1} \otimes I_{Nk})(I_N \otimes X) + (D_\Gamma^*)^{-1}]^{-1}$, and $\overline{\beta} = \overline{V}_\beta[(I_N \otimes X)^{'}(\Sigma^{-1} \otimes I_{Nk})y]$. The Full conditional posterior for Σ^{-1} is $W(\overline{S}^{-1}, \overline{\nu})$, with $\overline{S}^{-1} = (Y - XB)^{'}(Y - XB) + 2Q^{'}Q + \underline{S}^{-1}$ and $\overline{\nu} = T + 2Nk + \underline{\nu}$, with $vec(Q) = \Gamma$. The Full conditional posterior for λ_1^2 is $G(\bar{\mu}_{\lambda_1}, \bar{\nu}_{\lambda_1})$, where $\bar{\nu}_{\lambda_1} = \underline{\nu}_{\lambda_1} + 2N^2k$ and $\bar{\mu}_{\lambda_1} = \frac{\overline{\nu}_{\lambda_1}\underline{\mu}_{\lambda_1}}{\underline{\nu}_{\lambda_1} + 2\underline{\mu}_{\lambda_1}\sum \tau_j^2/\sigma_{\gamma_j}^2}$ Finally the full conditional posterior of $\frac{1}{2\eta_j^2}$ is Inverse Gaussian: $IG(\sqrt{\frac{\lambda_1^2}{\beta_j^2\sigma_{\gamma_j}^2}}, \frac{\lambda_1^2}{\sigma_{\gamma_j}^2})$.[1] Γ can not be directly drawn from the posteriors. But it can be recovered in each Gibbs iteration using the draws of $\frac{1}{2\eta_j^2}$ and Σ .

Conditional on arbitrary starting values, the Gibbs sampler contains the following six steps:

1. draw $\beta|\Sigma, \Lambda_1, \Gamma$ from $N(\overline{\beta}, \overline{V}_\beta)$;

2. draw $\Sigma^{-1}|\beta, \Lambda_1, \Gamma$ from $W(\overline{S}^{-1}, \overline{\nu})$

3. draw $\lambda_1^2|\Sigma, \beta, \Gamma$ from $G(\bar{\mu}_{\lambda_1}, \bar{\nu}_{\lambda_1})$

4. draw $\frac{1}{2\eta_j^2}|\beta, \Sigma, \Lambda_1$ from $IG(\sqrt{\frac{\lambda_1^2}{\beta_j^2\sigma_{\gamma_j}^2}}, \frac{\lambda_1^2}{\sigma_{\gamma_j}^2})$ for $j = 1, 2, ...N^2k$.

5. calculate Γ based on draws of Σ and $\frac{1}{2\eta_j^2}$ in the current iteration.

[1] We adopt the same form of the inverse-Gaussian density used in Park and Casella (2008).

1.2 Adaptive Lasso VAR Shrinkage

We define the adaptive Lasso estimator for a VAR as:

$$\hat{\beta}_{AL} = arg\ min_{\beta}\{[y-(I_n \otimes X)\beta]^{'}[y-(I_n \otimes X)\beta] + \sum_{j=1}^{N^2k} \lambda_{1,j}|\beta_j|\} \tag{12}$$

Correspondingly, the conditional multivariate mixture prior for β takes the following form:

$$\begin{aligned}\pi(\beta|\Sigma,\Gamma,\Lambda_1) \propto \prod_{j=1}^{N^2k} \{ \int_0^{\infty} \frac{1}{\sqrt{2\pi f_j(\Gamma))}} exp[-\frac{1}{2f_j(\Gamma)}\beta_j^2] d(f_j(\Gamma))\} \\ \times \{|M|^{-\frac{1}{2}} exp(-\frac{1}{2}\Gamma^{'} M^{-1}\Gamma)\}^2\end{aligned} \tag{13}$$

where $\Gamma = [\gamma_1, \gamma_2, ..., \gamma_{N^2k}]^{'}$, $M = \Sigma \otimes I_{Nk}$, and $f_j(\Gamma)$ is a function of Γ and Λ_1 to be defined later. In this mixture prior, the terms associated with the L_1 penalty are conditional on Σ through $f_j(\Gamma)$. In equation (13), the variances of β_a and β_b for $a \neq b$ are related through M. However, β_a and β_b themselves are independent of each other.

We need to find an appropriate $f_j(\Gamma)$ which provides us tractable posteriors. The last term in equation (13) takes the form of a multivariate Normal distribution $\Gamma \sim N(0, M)$. For ease of exposition, we first write the $N^2k \times N^2k$ covariance matrix M as following:

$$M = \begin{pmatrix} M_{1,1} & ... & M_{1,j} & M_{1,j+1} & ... & M_{1,N^2k} \\ ... & ... & ... & ... & ... & ... \\ M_{j,1} & ... & M_{j,j} & M_{j,j+1} & ... & M_{j,N^2k} \\ M_{j+1,1} & ... & M_{j+1,j} & M_{j+1,j+1} & ... & M_{j+1,N^2k} \\ ... & ... & ... & ... & ... & ... \\ M_{N^2k,1} & ... & M_{N^2k,j} & M_{N^2k,j+1} & ... & M_{N^2k,N^2k} \end{pmatrix} \tag{14}$$

Let $H_j = (M_{j,j+1}, ..., M_{j,N^2k}) \begin{pmatrix} M_{j+1,j+1} & ... & M_{j+1,N^2k} \\ ... & ... & ... \\ M_{N^2k,j+1} & ... & M_{N^2k,N^2k} \end{pmatrix}^{-1}$.

We next construct independent variables τ_j for $j = 1, 2, ..., N^2k$ using standard textbook techniques (e.g. Anderson, 2003; Muirhead 1982).

$$\tau_1 = \gamma_1 + H_1(\gamma_2, \gamma_3, ..., \gamma_{N^2k})' \tag{15}$$

$$\tau_2 = \gamma_2 + H_2(\gamma_3, \gamma_4, ..., \gamma_{N^2k})' \tag{16}$$

$$...$$

$$\tau_{N^2K-1} = \gamma_{N^2k-1} + H_{N^2k-1}\gamma_{N^2k} \tag{17}$$

$$\tau_{N^2K} = \gamma_{N^2k} \tag{18}$$

The joint density of $\tau_1, \tau_2, ..., \tau_{N^2k}$ is

$$N(\tau_1|0, \sigma^2_{\gamma_1})N(\tau_2|0, \sigma^2_{\gamma_2})...N(\tau_{N^2k}|0, \sigma^2_{\gamma_{N^2k}}) \tag{19}$$

where $\sigma^2_{\gamma_j} = M_{j,j} - H_j(M_{j,j+1}, ..., M_{j,N^2k})'$, with $\sigma^2_{\gamma_{N^2k}} = M_{N^2k,N^2k}$. Note that it is computationally feasible to derive $\sigma^2_{\gamma_j}$ when M is sparse.

The Jacobian of transforming $\Gamma \sim N(0, M)$ to (19) is 1. Defining $\eta_j = \tau_j/\lambda_{1,j}$, we can write (19) as

$$N(\eta_1|0, \sigma^2_{\gamma_1}\lambda^{-2}_{1,1})N(\eta_2|0, \sigma^2_{\gamma_2}\lambda^{-2}_{1,2})...N(\eta_{N^2k}|0, \sigma^2_{\gamma_{N^2k}}\lambda^{-2}_{1,N^2k}) \tag{20}$$

Let $f_j(\Gamma) = 2(\eta^2_j)$, the scale mixture prior is:

$$\begin{aligned}\pi(\beta|\Sigma, \Gamma, \Lambda_1) \propto \prod_{j=1}^{N^2k}\{ &\int_0^\infty \frac{1}{\sqrt{2\pi(2\eta^2_j))}} exp[-\frac{\beta^2_j}{2(2\eta^2_j)}]d(2\eta^2_j) \\ &\times \frac{\lambda^2_{1,j}}{2\sigma^2_{\gamma_j}} exp[-\frac{1}{2}\frac{2\eta^2_j}{(\sigma^2_{\gamma_j})/\lambda^2_{1,j}}]\}\end{aligned} \tag{21}$$

Equation (21) shows that the conditional prior for β_j is $N(0, \frac{1}{2\eta^2_j})$, and the conditional prior for β is

$$\beta|\Gamma, \Sigma, \Lambda_1, \Lambda_2 \sim N(0, D^*_\Gamma) \tag{22}$$

where $D^*_\Gamma = diag([\frac{1}{2\eta^2_1}, \frac{1}{2\eta^2_2}, ..., \frac{1}{2\eta^2_{N^2k}}])$.

Priors for Σ and $\lambda^2_{1,j}$ can be elicited following standard practice in VAR and Lasso literature. In this paper, we set Wishart prior for Σ^{-1} and Gamma prior for $\lambda^2_{1,j}$: $\Sigma^{-1} \sim W(\underline{S}^{-1}, \underline{\nu})$, $\lambda^2_{1,j} \sim G(\underline{\mu}_{\lambda^2_{1,j}}, \underline{\nu}_{\lambda^2_{1,j}})$.

The full conditional posterior for β is $\beta \sim N(\overline{\beta}, \overline{V}_\beta)$, where $\overline{V}_\beta = [(I_N \otimes X)')(\Sigma^{-1} \otimes I_{Nk})(I_N \otimes X) + (D^*_\Gamma)^{-1}]^{-1}$, and $\overline{\beta} = \overline{V}_\beta[(I_N \otimes X)'(\Sigma^{-1} \otimes I_{Nk})y]$. The Full conditional posterior for Σ^{-1} is $W(\overline{S}^{-1}, \overline{\nu})$, with $\overline{S}^{-1} = (Y - XB)'(Y - XB) + 2Q'Q + \underline{S}^{-1}$ and $\overline{\nu} = T + 2Nk + \underline{\nu}$, with $vec(Q) = \Gamma$. The

Full conditional posterior for $\lambda_{1,j}^2$ is $G(\bar{\mu}_{\lambda_{1,j}}, \bar{\nu}_{\lambda_{1,j}})$, where $\bar{\nu}_{\lambda_{1,j}} = \underline{\nu}_{\lambda_{1,j}} + 2$ and $\bar{\mu}_{\lambda_{1,j}} = \frac{\bar{\nu}_{\lambda_{1,j}} \sigma_j^2 \underline{\mu}_{\lambda_{1,j}}}{2\tau_j^2 \underline{\mu}_{\lambda_{1,j}} + \underline{\nu}_{\lambda_{1,j}} \sigma_{\gamma_j}^2}$. Finally the full conditional posterior of $\frac{1}{2\eta_j^2}$ is Inverse Gaussian: $IG(\sqrt{\frac{\lambda_{1,j}^2}{\beta_j^2 \sigma_{\gamma_j}^2}}, \frac{\lambda_{1,j}^2}{\sigma_{\gamma_j}^2})$. Γ can not be directly drawn from the posteriors. But it can be recovered in each Gibbs iteration using the draws of $\frac{1}{2\eta_j^2}$ and Σ .

Conditional on arbitrary starting values, the Gibbs sampler contains the following six steps:

1. draw $\beta|\Sigma, \Lambda_1, \Gamma$ from $N(\overline{\beta}, \overline{V}_\beta)$;

2. draw $\Sigma^{-1}|\beta, \Lambda_1, \Gamma$ from $W(\overline{S}^{-1}, \overline{\nu})$

3. draw $\lambda_{1,j}^2|\beta, \Sigma, \Lambda_{1,-j}, \Gamma$ from $G(\bar{\mu}_{\lambda_{1,j}}, \bar{\nu}_{\lambda_{1,j}})$ for $j = 1, 2, ...N^2k$

4. draw $\frac{1}{2\eta_j^2}|\beta, \Sigma, \Lambda_1$ from $IG(\sqrt{\frac{\lambda_{1,j}^2}{\beta_j^2 \sigma_{\gamma_j}^2}}, \frac{\lambda_{1,j}^2}{\sigma_{\gamma_j}^2})$ for $j = 1, 2, ...N^2k$.

5. calculate Γ based on draws of Σ and $\frac{1}{2\eta_j^2}$ in the current iteration.

1.3 E-net Lasso VAR Shrinkage

We define the e-net Lasso estimator for a VAR as:

$$\hat{\beta}_{EL} = arg\, min_\beta \{[y - (I_n \otimes X)\beta]^{'}[y - (I_n \otimes X)\beta] + \lambda_1 \sum_{j=1}^{N^2k} |\beta_j| + \lambda_2 \sum_{j=1}^{N^2k} \beta_j^2\} \tag{23}$$

Correspondingly, the conditional multivariate mixture prior for β takes the following form:

$$
\begin{aligned}
\pi(\beta|\Sigma,\Gamma,\lambda_1,\lambda_2) \propto \prod_{j=1}^{N^2k} \{ & \frac{\sqrt{\lambda_2}}{\sqrt{2\pi}} exp(-\frac{\lambda_2}{2}\beta_j^2) \\
& \times \int_0^\infty \frac{1}{\sqrt{2\pi f_j(\Gamma))}} exp[-\frac{1}{2f_j(\Gamma)}\beta_j^2] d(f_j(\Gamma))\} \\
& \times \{|M|^{-\frac{1}{2}} exp(-\frac{1}{2}\Gamma' M^{-1}\Gamma)\}^2
\end{aligned} \tag{24}
$$

where $\Gamma = [\gamma_1, \gamma_2, ..., \gamma_{N^2k}]'$, $M = \Sigma \otimes I_{Nk}$, and $f_j(\Gamma)$ is a function of Γ and Λ_1 to be defined later. In this mixture prior, the terms associated with the L_1 penalty are conditional on Σ through $f_j(\Gamma)$. In equation (24), the variances of β_a and β_b for $a \neq b$ are related through M. However, β_a and β_b themselves are independent of each other.

We need to find an appropriate $f_j(\Gamma)$ which provides us tractable posteriors. The last term in equation (24) takes the form of a multivariate Normal distribution $\Gamma \sim N(0, M)$. For ease of exposition, we first write the $N^2k \times N^2k$ covariance matrix M as following:

$$
M = \begin{pmatrix}
M_{1,1} & ... & M_{1,j} & M_{1,j+1} & ... & M_{1,N^2k} \\
... & ... & ... & ... & ... & ... \\
M_{j,1} & ... & M_{j,j} & M_{j,j+1} & ... & M_{j,N^2k} \\
M_{j+1,1} & ... & M_{j+1,j} & M_{j+1,j+1} & ... & M_{j+1,N^2k} \\
... & ... & ... & ... & ... & ... \\
M_{N^2k,1} & ... & M_{N^2k,j} & M_{N^2k,j+1} & ... & M_{N^2k,N^2k}
\end{pmatrix} \tag{25}
$$

Let $H_j = (M_{j,j+1}, ..., M_{j,N^2k}) \begin{pmatrix} M_{j+1,j+1} & ... & M_{j+1,N^2k} \\ ... & ... & ... \\ M_{N^2k,j+1} & ... & M_{N^2k,N^2k} \end{pmatrix}^{-1}$.

We next construct independent variables τ_j for $j = 1, 2, ..., N^2k$ using standard textbook techniques (e.g. Anderson, 2003; Muirhead 1982).

$$\tau_1 = \gamma_1 + H_1(\gamma_2, \gamma_3, ..., \gamma_{N^2k})' \tag{26}$$

$$\tau_2 = \gamma_2 + H_2(\gamma_3, \gamma_4, ..., \gamma_{N^2k})' \tag{27}$$

$$...$$

$$\tau_{N^2K-1} = \gamma_{N^2k-1} + H_{N^2k-1}\gamma_{N^2k} \tag{28}$$

$$\tau_{N^2K} = \gamma_{N^2k} \tag{29}$$

The joint density of $\tau_1, \tau_2, ..., \tau_{N^2k}$ is

$$N(\tau_1|0, \sigma^2_{\gamma_1})N(\tau_2|0, \sigma^2_{\gamma_2})...N(\tau_{N^2k}|0, \sigma^2_{\gamma_{N^2k}}) \tag{30}$$

where $\sigma^2_{\gamma_j} = M_{j,j} - H_j(M_{j,j+1}, ..., M_{j,N^2k})'$, with $\sigma^2_{\gamma_{N^2k}} = M_{N^2k,N^2k}$. Note that it is computationally feasible to derive $\sigma^2_{\gamma_j}$ when M is sparse.

The Jacobian of transforming $\Gamma \sim N(0, M)$ to (30) is 1. Defining $\eta_j = \tau_j/\lambda_1$, we can write (30) as

$$N(\eta_1|0, \sigma^2_{\gamma_1}\lambda_1^{-2})N(\eta_2|0, \sigma^2_{\gamma_2}\lambda_1^{-2})...N(\eta_{N^2k}|0, \sigma^2_{\gamma_{N^2k}}\lambda_1^{-2}) \tag{31}$$

Let $f_j(\Gamma) = 2(\eta_j^2)$, the scale mixture prior is:

$$\pi(\beta|\Sigma, \Gamma, \lambda_1, \lambda_2) \propto \prod_{j=1}^{N^2 k} \{ \frac{\sqrt{\lambda_2}}{\sqrt{2\pi}} exp(-\frac{\lambda_2}{2}\beta_j^2) \\ \times \int_0^\infty \frac{1}{\sqrt{2\pi(2\eta_j^2))}} exp[-\frac{\beta_j^2}{2(2\eta_j^2)}] d(2\eta_j^2) \\ \times \frac{\lambda_1^2}{2\sigma_{\gamma_j}^2} exp[-\frac{1}{2}\frac{2\eta_j^2}{(\sigma_{\gamma_j}^2)/\lambda_1^2}] \} \tag{32}$$

where $\eta_j = \tau_j/\lambda_1$.

The last two terms in (32) constitute a scale mixture of Normals (with an exponential mixing density), which can be expressed as the univariate Laplace distribution $\frac{\lambda_1}{2\sqrt{\sigma_{\gamma_j}^2}} exp(-\frac{\lambda_1}{\sqrt{\sigma_{\gamma_j}^2}}|\beta_j|)$.

Equation (32) shows that the conditional prior for β_j is $N(0, \frac{2\eta_j^2}{2\lambda_2\eta_j^2+1})$, and the conditional prior for β is

$$\beta|\Gamma, \Sigma, \Lambda_1, \Lambda_2 \sim N(0, D_\Gamma^*) \tag{33}$$

where $D_\Gamma^* = diag([\frac{2\eta_1^2}{2\lambda_2\eta_1^2+1}, \frac{2\eta_2^2}{2\lambda_2\eta_2^2+1}, ..., \frac{2\eta_{N^2k}^2}{2\lambda_2\eta_{N^2k}^2+1}])$.

Priors for Σ and λ_1^2 can be elicited following standard practice in VAR and Lasso literature. In this paper, we set Wishart prior for Σ^{-1} and Gamma priors for λ_1^2 and λ_2: $\Sigma^{-1} \sim W(\underline{S}^{-1}, \underline{\nu})$, $\lambda_1^2 \sim G(\underline{\mu}_{\lambda_1^2}, \underline{\nu}_{\lambda_1^2})$, $\lambda_2 \sim G(\underline{\mu}_{\lambda_2}, \underline{\nu}_{\lambda_2})$.

The full conditional posterior for β is $\beta \sim N(\overline{\beta}, \overline{V}_\beta)$, where $\overline{V}_\beta = [(I_N \otimes X)^{'})(\Sigma^{-1} \otimes I_{Nk})(I_N \otimes X) + (D_\Gamma^*)^{-1}]^{-1}$, and $\overline{\beta} = \overline{V}_\beta[(I_N \otimes X)^{'}(\Sigma^{-1} \otimes I_{Nk})y]$. The Full conditional posterior for Σ^{-1} is $W(\overline{S}^{-1}, \overline{\nu})$, with $\overline{S}^{-1} = (Y - XB)^{'}(Y - XB) + 2Q^{'}Q + \underline{S}^{-1}$ and $\overline{\nu} = T + 2Nk + \underline{\nu}$, with $vec(Q) = \Gamma$. The

Full conditional posterior for λ_1^2 is $G(\bar{\mu}_{\lambda_1}, \bar{\nu}_{\lambda_1})$, where $\bar{\nu}_{\lambda_1} = \underline{\nu}_{\lambda_1} + 2N^2k$ and $\bar{\mu}_{\lambda_1} = \frac{\bar{\nu}_{\lambda_1}\underline{\mu}_{\lambda_1}}{\underline{\nu}_{\lambda_1} + 2\underline{\mu}_{\lambda_1}\sum \tau_j^2/\sigma_{\gamma_j}^2}$. The Full conditional posterior for λ_2 is $G(\bar{\mu}_{\lambda_2}, \bar{\nu}_{\lambda_2})$, where $\bar{\nu}_{\lambda_2} = \underline{\nu}_{\lambda_2} + N^2k$ and $\bar{\mu}_{\lambda_2} = \frac{\underline{\mu}_{\lambda_2}\bar{\nu}_{\lambda_2}}{\underline{\nu}_{\lambda_2} + \underline{\mu}_{\lambda_2}\sum \beta_j^2}$. Finally the full conditional posterior of $\frac{1}{2\eta_j^2}$ is Inverse Gaussian: $IG(\sqrt{\frac{\lambda_1^2}{\beta_j^2\sigma_{\gamma_j}^2}}, \frac{\lambda_1^2}{\sigma_{\gamma_j}^2})$. Γ can not be directly drawn from the posteriors. But it can be recovered in each Gibbs iteration using the draws of $\frac{1}{2\eta_j^2}$ and Σ .

Conditional on arbitrary starting values, the Gibbs sampler contains the following six steps:

1. draw $\beta|\Sigma, \Lambda_1, \Lambda_2, \Gamma$ from $N(\overline{\beta}, \overline{V}_\beta)$;
2. draw $\Sigma^{-1}|\beta, \Lambda_1, \Lambda_2, \Gamma$ from $W(\overline{S}^{-1}, \overline{\nu})$
3. draw $\lambda_1^2|\beta, \Sigma, \Lambda_2, \Gamma$ from $G(\bar{\mu}_{\lambda_1}, \bar{\nu}_{\lambda_1})$
4. draw $\lambda_2|\beta, \Sigma, \Lambda_1, \Gamma$ from $G(\bar{\mu}_{\lambda_2}, \bar{\nu}_{\lambda_2})$
5. draw $\frac{1}{2\eta_j^2}|\beta, \Sigma, \Lambda_1, \Lambda_2$ from $IG(\sqrt{\frac{\lambda_1^2}{\beta_j^2\sigma_{\gamma_j}^2}}, \frac{\lambda_1^2}{\sigma_{\gamma_j}^2})$ for $j = 1, 2, ...N^2k$.
6. calculate Γ based on draws of Σ and $\frac{1}{2\eta_j^2}$ in the current iteration.

1.4 Adaptive E-net Lasso VAR Shrinkage

In line with Zou and Zhang (2009), we define the adaptive e-net Lasso estimator for a VAR as following:

$$\hat{\beta}_{AEL} = arg\ min_\beta \{[y - (I_n \otimes X)\beta]'[y - (I_n \otimes X)\beta] + \sum_{j=1}^{N^2k} \lambda_{1,j}|\beta_j| + \lambda_2 \sum_{j=1}^{N^2k} \beta_j^2\} \tag{34}$$

Correspondingly, the conditional multivariate mixture prior for β takes the following form:

$$\begin{aligned}\pi(\beta|\Sigma,\Gamma,\Lambda_1,\lambda_2) \propto \prod_{j=1}^{N^2k} &\{ \frac{\sqrt{\lambda_2}}{\sqrt{2\pi}} exp(-\frac{\lambda_2}{2}\beta_j^2) \\ &\times \int_0^\infty \frac{1}{\sqrt{2\pi f_j(\Gamma))}} exp[-\frac{1}{2f_j(\Gamma)}\beta_j^2] d(f_j(\Gamma))\} \\ &\times \{|M|^{-\frac{1}{2}} exp(-\frac{1}{2}\Gamma' M^{-1}\Gamma)\}^2\end{aligned} \tag{35}$$

where $\Gamma = [\gamma_1, \gamma_2, ..., \gamma_{N^2k}]'$, $M = \Sigma \otimes I_{Nk}$, and $f_j(\Gamma)$ is a function of Γ and Λ_1 to be defined later. In this mixture prior, the terms associated with the L_1 penalty are conditional on Σ through $f_j(\Gamma)$.

We need to find an appropriate $f_j(\Gamma)$ which provides us tractable posteriors. The last term in equation (35) takes the form of a multivariate Normal distribution $\Gamma \sim N(0, M)$. For ease of exposition, we first write the $N^2k \times N^2k$ covariance matrix M as following:

$$M = \begin{pmatrix} M_{1,1} & ... & M_{1,j} & M_{1,j+1} & ... & M_{1,N^2k} \\ ... & ... & ... & ... & ... & ... \\ M_{j,1} & ... & M_{j,j} & M_{j,j+1} & ... & M_{j,N^2k} \\ M_{j+1,1} & ... & M_{j+1,j} & M_{j+1,j+1} & ... & M_{j+1,N^2k} \\ ... & ... & ... & ... & ... & ... \\ M_{N^2k,1} & ... & M_{N^2k,j} & M_{N^2k,j+1} & ... & M_{N^2k,N^2k} \end{pmatrix} \tag{36}$$

Let $H_j = (M_{j,j+1}, ..., M_{j,N^2k}) \begin{pmatrix} M_{j+1,j+1} & ... & M_{j+1,N^2k} \\ ... & ... & ... \\ M_{N^2k,j+1} & ... & M_{N^2k,N^2k} \end{pmatrix}^{-1}$.

We next construct independent variables τ_j for $j = 1, 2, ..., N^2k$ using standard textbook techniques (e.g. Anderson, 2003; Muirhead 1982).

$$\tau_1 = \gamma_1 + H_1(\gamma_2, \gamma_3, ..., \gamma_{N^2k})' \tag{37}$$

$$\tau_2 = \gamma_2 + H_2(\gamma_3, \gamma_4, ..., \gamma_{N^2k})' \tag{38}$$

$$...$$

$$\tau_{N^2K-1} = \gamma_{N^2k-1} + H_{N^2k-1}\gamma_{N^2k} \tag{39}$$

$$\tau_{N^2K} = \gamma_{N^2k} \tag{40}$$

The joint density of $\tau_1, \tau_2, ..., \tau_{N^2k}$ is

$$N(\tau_1|0, \sigma^2_{\gamma_1})N(\tau_2|0, \sigma^2_{\gamma_2})...N(\tau_{N^2k}|0, \sigma^2_{\gamma_{N^2k}}) \tag{41}$$

where $\sigma^2_{\gamma_j} = M_{j,j} - H_j(M_{j,j+1}, ..., M_{j,N^2k})'$, with $\sigma^2_{\gamma_{N^2k}} = M_{N^2k,N^2k}$. Note that it is computationally feasible to derive $\sigma^2_{\gamma_j}$ when M is sparse.

The Jacobian of transforming $\Gamma \sim N(0, M)$ to (41) is 1. Defining $\eta_j = \tau_j/\lambda_{1,j}$, we can write (41) as

$$N(\eta_1|0, \sigma^2_{\gamma_1}\lambda^{-2}_{1,1})N(\eta_2|0, \sigma^2_{\gamma_2}\lambda^{-2}_{1,2})...N(\eta_{N^2k}|0, \sigma^2_{\gamma_{N^2k}}\lambda^{-2}_{1,N^2k}) \tag{42}$$

Let $f_j(\Gamma) = 2(\eta_j^2)$. The scale mixture prior in (35) can be rewritten as:

$$\begin{aligned}\pi(\beta|\Sigma,\Gamma,\Lambda_1,\lambda_2) \propto \prod_{j=1}^{N^2k} \{ &\frac{\sqrt{\lambda_2}}{\sqrt{2\pi}} exp(-\frac{\lambda_2}{2}\beta_j^2) \\ &\times \int_0^\infty \frac{1}{\sqrt{2\pi(2\eta_j^2)}} exp[-\frac{\beta_j^2}{2(2\eta_j^2)}] d(2\eta_j^2) \\ &\times \frac{\lambda_{1,j}^2}{2\sigma_{\gamma_j}^2} exp[-\frac{1}{2}\frac{2\eta_j^2}{(\sigma_{\gamma_j}^2)/\lambda_{1,j}^2}] \}\end{aligned} \tag{43}$$

The last two terms in (43) constitute a scale mixture of Normals (with an exponential mixing density), which can be expressed as the univariate Laplace distribution $\frac{\lambda_{1,j}}{2\sqrt{\sigma_{\gamma_j}^2}} exp(-\frac{\lambda_{1,j}}{\sqrt{\sigma_{\gamma_j}^2}}|\beta_j|)$.

Equation (43) shows that the conditional prior for β_j is $N(0, \frac{2\eta_j^2}{2\lambda_2\eta_j^2+1})$, and the conditional prior for β is

$$\beta|\Gamma,\Sigma,\Lambda_1,\Lambda_2 \sim N(0, D_\Gamma^*) \tag{44}$$

where $D_\Gamma^* = diag([\frac{2\eta_1^2}{2\lambda_2\eta_1^2+1}, \frac{2\eta_2^2}{2\lambda_2\eta_2^2+1}, ..., \frac{2\eta_{N^2k}^2}{2\lambda_2\eta_{N^2k}^2+1}])$.

Priors for Σ and $\lambda_{1,j}^2$ can be elicited following standard practice in VAR and Lasso literature. In this paper, we set Wishart prior for Σ^{-1} and Gamma priors for $\lambda_{1,j}^2$ and $\lambda_{2,j}$: $\Sigma^{-1} \sim W(\underline{S}^{-1}, \underline{\nu})$, $\lambda_{1,j}^2 \sim G(\underline{\mu}_{\lambda_{1,j}^2}, \underline{\nu}_{\lambda_{1,j}^2})$, $\lambda_{2,j} \sim G(\underline{\mu}_{\lambda_2}, \underline{\nu}_{\lambda_2})$.

The full conditional posterior for β is $\beta \sim N(\overline{\beta}, \overline{V}_\beta)$, where $\overline{V}_\beta = [(I_N \otimes X)^{'})(\Sigma^{-1} \otimes I_{Nk})(I_N \otimes X) + (D_\Gamma^*)^{-1}]^{-1}$, and $\overline{\beta} = \overline{V}_\beta[(I_N \otimes X)^{'}(\Sigma^{-1} \otimes I_{Nk})y]$. The Full conditional posterior for Σ^{-1} is $W(\overline{S}^{-1}, \overline{\nu})$, with $\overline{S}^{-1} = (Y - XB)^{'}(Y - XB) + 2Q^{'}Q + \underline{S}^{-1}$ and $\overline{\nu} = T + 2Nk + \underline{\nu}$, with $vec(Q) = \Gamma$. The

Full conditional posterior for $\lambda_{1,j}^2$ is $G(\bar{\mu}_{\lambda_{1,j}}, \bar{\nu}_{\lambda_{1,j}})$, where $\bar{\nu}_{\lambda_{1,j}} = \underline{\nu}_{\lambda_{1,j}} + 2$ and $\bar{\mu}_{\lambda_{1,j}} = \frac{\bar{\nu}_{\lambda_{1,j}} \sigma_j^2 \underline{\mu}_{\lambda_{1,j}}}{2\tau_j^2 \underline{\mu}_{\lambda_{1,j}} + \underline{\nu}_{\lambda_{1,j}} \sigma_{\gamma_j}^2}$. The Full conditional posterior for λ_2 is $G(\bar{\mu}_{\lambda_2}, \bar{\nu}_{\lambda_2})$, where $\bar{\nu}_{\lambda_2} = \underline{\nu}_{\lambda_2} + N^2 k$ and $\bar{\mu}_{\lambda_2} = \frac{\underline{\mu}_{\lambda_2} \bar{\nu}_{\lambda_2}}{\underline{\nu}_{\lambda_2} + \underline{\mu}_{\lambda_2} \sum \beta_j^2}$. Finally the full conditional posterior of $\frac{1}{2\eta_j^2}$ is Inverse Gaussian: $IG(\sqrt{\frac{\lambda_{1,j}^2}{\beta_j^2 \sigma_{\gamma_j}^2}}, \frac{\lambda_{1,j}^2}{\sigma_{\gamma_j}^2})$. Γ can not be directly drawn from the posteriors. But it can be recovered in each Gibbs iteration using the draws of $\frac{1}{2\eta_j^2}$ and Σ .

Conditional on arbitrary starting values, the Gibbs sampler contains the following six steps:

1. draw $\beta|\Sigma, \Lambda_1, \Lambda_2, \Gamma$ from $N(\overline{\beta}, \overline{V}_\beta)$;
2. draw $\Sigma^{-1}|\beta, \Lambda_1, \Lambda_2, \Gamma$ from $W(\overline{S}^{-1}, \overline{\nu})$
3. draw $\lambda_{1,j}^2|\beta, \Sigma, \Lambda_{1,-j}, \Lambda_2, \Gamma$ from $G(\bar{\mu}_{\lambda_{1,j}}, \bar{\nu}_{\lambda_{1,j}})$ for $j = 1, 2, ...N^2 k$
4. draw $\lambda_2|\beta, \Sigma, \Lambda_1, \Gamma$ from $G(\bar{\mu}_{\lambda_2}, \bar{\nu}_{\lambda_2})$
5. draw $\frac{1}{2\eta_j^2}|\beta, \Sigma, \Lambda_1, \Lambda_2$ from $IG(\sqrt{\frac{\lambda_{1,j}^2}{\beta_j^2 \sigma_{\gamma_j}^2}}, \frac{\lambda_{1,j}^2}{\sigma_{\gamma_j}^2})$ for $j = 1, 2, ...N^2 k$.
6. calculate Γ based on draws of Σ and $\frac{1}{2\eta_j^2}$ in the current iteration.

2 Detailed Forecast Evaluation Results

Tables 1-4 report the DAELasso forecasts results along with Lasso, adaptive Lasso, e-net Lasso, adaptive e-net Lasso, and those of the factor models and the seven popular Bayesian shrinkage priors in Koop (2011). In line with Koop (2011), we present MSFE relative to the random walk and log predictive likelihood for GDP, CPI and FFR. The results for DAELasso and four other Lasso types of shrinkage methods are reported at the top of each table, followed by those of the methods reported in Koop (2011). Koop (2011) considers three variants of the Minnesota prior. The first is the natural conjugate prior used in Banbura et al (2010), which is labelled 'Minn. Prior as in BGR'. The second is the traditional Minnesota prior of Litterman (1986), which is labelled 'Minn. Prior Σ diagonal'. The third is the traditional Minnesota prior except that the upper left 3×3 block of Σ is not assumed to be daigonal, which is labelled 'Minn. Prior Σ not diagonal'. Koop (2011) also evaluates the performances of four types of SSVS priors. The first is the same as George et al (2008), which is labelled 'SSVS Non-conj. semi-automatic'. The second is a combination of the non-conjugate SSVS prior and Minnesota prior with variables selected in a data based fashion, which is labelled 'SSVS Non-conj. plus Minn. Prior'. The Third is a conjugate SSVS prior, which is labelled 'SSVS Conjugate Semi-automatic'. The fourth is a combination of the conjugate SSVS prior and Minnesota prior, which is labelled 'SSVS Conjugate plus Minn. Prior'. Finally the results for factor-augmented VAR models with one and four lagged factors are labelled as 'Factor model p=1' and 'Factor model p=4', respectively. We refer to Koop (2011) for a lucid description of these priors.

Table 1: Rolling Forecasting for $h = 1$

	GDP	CPI	FFR
DAELasso	0.58 (-198.9)	0.32 (-192.7)	0.57 (-211.7)
adaptive e-net Lasso	0.67 (-195.8)	0.40 (-199.4)	0.63 (-215.0)
e-net Lasso	0.68 (-215.3)	0.40 (-211.6)	0.63 (-223.7)
adaptive Lasso	0.77 (-225.6)	0.31 (-209.2)	0.62 (-228.3)
Lasso	0.67 (-255.8)	0.39 (-241.3)	0.63 (-257.6)
Minn. Prior as in BGR	0.58 (-190.5)	0.34 (-209.2)	0.51 (-177.4)
Minn. Prior Σ diagonal	0.61 (-194.0)	0.30 (-193.0)	0.52 (-181.7)
Minn. Prior Σ not diagonal	0.61 (-192.1)	0.31 (-202.4)	0.53 (-185.9)
SSVS Conjugate semi-automatic	0.81 (-209.4)	0.38 (-231.8)	0.63 (-175.8)
SSVS Conjugate plus Minn. Prior	0.59 (-191.4)	0.35 (-212.1)	0.51 (-179.2)
SSVS Non-conj. semi-automatic	0.88 (-234.3)	0.47 (-236.0)	0.73 (-213.0)
SSVS Non-conj. plus Minn. Prior	0.68 (-197.9)	0.34 (-195.2)	0.52 (-177.2)
Factor model p=1	1.21 (-252.8)	0.59 (-242.7)	1.42 (-236.4)
Factor model p=4	4.46 (-401.7)	1.88 (-457.0)	2.88 (-352.7)

Notes:

MSFEs as proportion of random walk MSFEs.

Sum of log predictive likelihoods in parentheses.

Table 2: Rolling Forecasting for $h = 4$

	GDP	CPI	FFR
DAELasso	0.55 (-206.9)	0.48 (-205.9)	0.65 (-230.9)
adaptive e-net Lasso	0.53 (-195.7)	0.47 (-204.4)	0.55 (-219.9)
e-net Lasso	0.53 (-215.2)	0.47 (-213.5)	0.55 (-225.5)
adaptive Lasso	0.74 (-233.9)	0.54 (-223.0)	0.78 (-247.7)
Lasso	0.53 (-255.9)	0.47 (-242.6)	0.55 (-259.0)
Minn. Prior as in BGR	0.59 (-217.1)	0.55 (-227.7)	0.59 (-213.4)
Minn. Prior Σ diagonal	0.59 (-211.1)	0.55 (-232.4)	0.59 (-246.6)
Minn. Prior Σ not diagonal	0.58 (-210.6)	0.58 (-222.2)	0.58 (-212.1)
SSVS Conjugate semi-automatic	1.23 (-282.6)	0.99 (-284.3)	1.32 (-273.8)
SSVS Conjugate plus Minn. Prior	0.63 (-230.2)	0.54 (-221.2)	0.61 (-213.5)
SSVS Non-conj. semi-automatic	1.60 (-294.1)	1.22 (-266.2)	1.64 (-268.8)
SSVS Non-conj. plus Minn. Prior	0.63 (-209.9)	0.51 (-201.3)	0.58 (-198.1)
Factor model p=1	1.39 (-280.1)	0.91 (-255.5)	1.35 (-283.4)
Factor model p=4	5.03 (-562.9)	3.64 (-522.3)	6.73 (-593.8)

Notes:

MSFEs as proportion of random walk MSFEs.

Sum of log predictive likelihoods in parentheses.

Table 3: Recursive Forecasting for $h = 1$

	GDP	CPI	FFR
DAELasso	0.55 (-210.4)	0.29 (-190.9)	0.56 (-224.2)
adaptive e-net Lasso	0.67 (-242.0)	0.40 (-201.6)	0.63 (-239.8)
e-net Lasso	0.68 (-225.6)	0.40 (-212.5)	0.63 (-237.9)
adaptive Lasso	0.62 (-219.2)	0.28 (-196.4)	0.60 (-226.8)
Lasso	0.67 (-236.5)	0.39 (-221.1)	0.62 (-242.9)
Minn. Prior as in BGR	0.56 (-192.3)	0.30 (-195.9)	0.51 (-229.1)
Minn. Prior Σ diagonal	0.58 (-204.3)	0.28 (-182.2)	0.54 (-238.8)
Minn. Prior Σ not diagonal	0.55 (-195.4)	0.27 (-184.1)	0.52 (-249.5)
SSVS Conjugate semi-automatic	0.68 (-199.9)	0.27 (-191.2)	0.63 (-245.3)
SSVS Conjugate plus Minn. Prior	0.56 (-192.5)	0.31 (-197.6)	0.51 (-228.5)
SSVS Non-conj. semi-automatic	0.64 (-205.1)	0.32 (-196.5)	0.58 (-237.2)
SSVS Non-conj. plus Minn. Prior	0.65 (-203.9)	0.29 (-187.6)	0.54 (-228.9)
Factor model p=1	0.68 (-198.3)	0.30 (-193.2)	0.67 (-227.9)
Factor model p=4	0.90 (-212.9)	0.35 (-219.1)	0.77 (-245.6)

Notes:

MSFEs as proportion of random walk MSFEs.

Sum of log predictive likelihoods in parentheses.

Table 4: Recursive Forecasting for $h = 4$

	GDP	CPI	FFR
DAELasso	0.54 (-218.3)	0.48 (-206.6)	0.61 (-239.6)
adaptive e-net Lasso	0.53 (-215.6)	0.47 (-207.0)	0.55 (-247.3)
e-net Lasso	0.53 (-225.5)	0.47 (-213.7)	0.55 (-239.0)
adaptive Lasso	0.63 (-228.0)	0.52 (-214.7)	0.66 (-242.2)
Lasso	0.53 (-236.2)	0.47 (-222.8)	0.55 (-244.3)
Minn. Prior as in BGR	0.61 (-214.7)	0.52 (-219.4)	0.59 (-249.6)
Minn. Prior Σ diagonal	0.61 (-214.0)	0.52 (-217.6)	0.61 (-278.1)
Minn. Prior Σ not diagonal	0.62 (-213.3)	0.52 (-216.1)	0.59 (-244.8)
SSVS Conjugate semi-automatic	0.65 (-212.4)	0.60 (-225.0)	0.59 (-249.5)
SSVS Conjugate plus Minn. Prior	0.84 (-219.6)	0.70 (-246.6)	0.67 (-258.5)
SSVS Non-conj. semi-automatic	0.75 (-293.2)	0.77 (-226.4)	0.88 (-268.1)
SSVS Non-conj. plus Minn. Prior	0.67 (-219.0)	0.49 (-201.6)	0.53 (-233.7)
Factor model p=1	0.84 (-228.9)	0.55 (-211.6)	0.69 (-244.1)
Factor model p=4	0.89 (-243.6)	0.62 (-227.4)	0.68 (-249.1)

Notes:

MSFEs as proportion of random walk MSFEs.

Sum of log predictive likelihoods in parentheses.