Bayesian Doubly Adaptive Elastic-Net Lasso For VAR Shrinkage

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

January 10, 2012

Abstract

We develop a novel Bayesian doubly adaptive elastic-net Lasso (DAELasso) approach for VAR shrinkage. DAELasso achieves data selection and coefficients shrinkage in a data based manner. It constructively deals with the explanatory variables that tend to be highly collinear by encouraging grouping effect. In addition, it allows for different degree of shrinkages for different coefficients. Rewriting the multivariate Laplace distribution as a scale mixture, we establish closed-form posteriors that can be drawn from a Gibbs sampler. We compare the forecasting performance of DAELasso to that of other popular Bayesian methods using US macro economic data. The results suggest that DAELasso is a useful complement to the available Baysian VAR shrinkage methods.

*The author would like to thank Gary Koop and participants of CFE11 conference in London for their constructive comments. Any remaining errors are the author's responsibility.

1 Introduction

Since the seminal work of Sims (1972, 1980), vector autoregressive (VAR) models have been widely used for analyzing the interrelationship between economic series and forecasting. VAR models are parameter rich and subject to the curse of dimensionality. With the increasing availability of data, it is not uncommon for researchers to have a VAR with the number of coefficients exceeding the number of observations. In this case, the coefficients can no longer be uniquely estimated. Various Bayesian shrinkage methods have proved successful in circumventing this problem by imposing appropriate prior restrictions. Following the notion of Chudik and Pesaran (2011), we can divide the popular Bayesian VAR shrinkage approaches into three categories: (i) shrinkage of the parameter space, (ii) shrinkage of the variable space, and (iii) shrinkage of both the parameter space and variable space. Traditional Minnesota prior of Doan et al (1984) and Litterman (1986) and its natural variants (e.g. Kadiyala and Karlsson, 1997; Banbura et al, 2010), shrink the parameters through priors controlling the degree of shrinkage. The stochastic search variable selection (SSVS) prior of George et al (2008) shrinks variable space by including/excluding different explanatory variables. Finally, the family of SSVS plus Minnesota priors of Koop (2011) combine variable selection and parameter shrinkage.

Despite being successful, these Bayesian VAR shrinkage methods have their own drawbacks. A main feature of Minnesota priors is that they discriminate between the parameters on own lags and that on lags of other variables. By doing so, Minnesota priors risk diminishing the effects of

possible leading indicators and coincident variables, whose importance for forecasting has been intensively investigated since first being advocated by Burns and Mitchell at the National Bureau of Economic Research in 1937 (e.g. Burns and Mitchell, 1946; Stock and Watson, 1989). By contrast, SSVS prior for VAR is not subject to the criticism of being subjective as it relies on data information to select variables. Yet, SSVS VAR prior is developed from its single equation counterpart of George and McCulloch (1993, 1997), and in the single equation set-up, standard SSVS prior is known for tending to select highly correlated variables at the cost of ignoring others that may improve the model's forecasting performance (Kwon et al, 2011).

Tibshirani's (1996) least absolute shrinkage and selection operator (Lasso) and its variants, such as fused Lasso of Tibshirani et al (2005), elastic net Lasso (e-net Lasso) of Zou and Hastie (2005), and grouped Lasso of Yuan and Lin (2007), offer the potential for alternative approaches to VAR shrinkage. Lasso methods are attractive as they can select variables and shrink parameters in a data based manner. Recently, Bayesian Lasso has gained popularity as it can be easily implemented through MCMC or a Gibbs sampler (e.g. Park and Casella, 2008; Kyung et al, 2010), and it can automatically achieve adaptive shrinkage to allow for different degree of shrinkage (e.g. Griffin and Brown, 2010; Leng et al, 2010). Despite being successful, the Lasso literature is mainly concentrated on single equation models. To our best knowledge, only a few studies in the frequentist framework explore Lasso shrinkage for VARs (e.g., Hsu, Hung and Chang, 2008; Song and Bickel, 2011). And these available methods can be too restrictive as they either assume the covariance matrix of the VAR errors to be diagonal or

assume its off-diagonal elements are much smaller than the diagonal ones.

This paper develops a novel Bayesian Lasso method for VAR shrinkage: the doubly adaptive e-net Lasso (DAELasso) that suitable for VAR models in macroeconomic research. Considering VAR models usually have highly correlated explanatory variables and sometime have more coefficients than the observations, we use the e-net Lasso of Zou and Hastie (2005) to deal with the problem of multicollinearity. While Lasso generally only picks up one variable among a group of highly correlated variables, e-net Lasso has the potential of selecting all the important variables regardless the number of observations (Zou and Hastie, 2005). Note that if we have formal and informal economic theory at hand to group the data, it can be more desirable to have other type of Lasso, such as the fused Lasso and grouped Lasso, instead of e-net Lasso for VAR shrinkage. Yet in general we do not have such information, and e-net Lasso turns out to be the most appealing choice for it encourages the grouping effect supported by the data itself. Finally, considering that the traditional e-net Lasso can give the wrong results as it imposes the same amount of shrinkage for all the coefficients (Zou and Zhang, 2009), we extend the Bayesian adaptive Lasso of Leng et al (2010) to e-net Lasso. We call the approach 'doubly adaptive' to reflect the fact that unlike the adaptive e-net Lasso of Zou and Zhang (2009) which only adapts the tuning parameters of the L_1 norm, we allow for tuning parameters of both the L_1 and L_2 norms to be adapted. DAELasso is flexible, but it can be too complicated for some data. Thus, in this paper we also introduce four alternative Lasso types of VAR shrinkage methods: Lasso, adaptive Lasso, e-net Lasso, and adaptive e-net Lasso, each of them is a nested version of

DAELasso.

In empirical work, we evaluate the forecasting performance of DAELasso approach alongside Lasso, adaptive Lasso, e-net Lasso, and adaptive e-net Lasso. In addition, we compare the forecasting performance of these Lasso types of methods with that of other seven popular Baysian VAR shrinkage methods reviewed in Koop (2011). We employ Koop's (2011) data set that contains 20 US macroeconomic series, which is originally compiled by James H. Stock and Mark W. Watson. The data runs from 1959Q1 to 2008Q4. In line with Koop (2011), we conduct rolling and recursive forecast exercises and calculate both the mean squared forecast error (MSFE) and predictive likelihood measures. Using relatively uninformative priors, we find DAELasso approach leads to forecasting results that compares favorably or equally well to other Bayesian VAR shrinkage methods. This suggests that DAELasso approach is an appropriate complement to the available Baysian VAR toolkit.

The remainder of the paper is organized as following. Section 2 develops the Bayesian DAELasso methods. Section 3 presents four alternative Lasso types of VAR shrinkage methods that nested in DAELasso. Section 4 compares the forecasting performance of Lasso methods and the seven popular Bayesian VAR shrinkage methods reviewed in Koop (2011). Section 5 concludes. A data list is provided in the appendix.

2 The Estimator and Bayesian Methods

Without loss of generality, we assume that all the variables are centered. Let Y be a $T \times N$ dependent variables, X be a $T \times Nk$ matrix contains the k lags of each dependent variable, and B be the coefficient matrix of dimension $Nk \times N$. In matrix form, an unrestricted VAR model of N variables takes the following form:

$$Y = XB + E \tag{1}$$

where E is a $T \times N$ matrix for *i.i.d.* error terms with its t^{th} row distributed as $\mathrm{N}(0, \Sigma)$.

Given the assumptions of the error term, the likelihood function of model (1) can be expressed as

$$L(b, \Sigma) \propto |\Sigma|^{-\frac{T}{2}} exp\{-\frac{1}{2} tr(Y - XB)^{'}(Y - XB)\Sigma^{-1}\} \tag{2}$$

Note that when $X^{'}X$ is not of full rank, which is often the case when we have more parameters than the number of observations, the least squares estimator $B_{LS} = (X^{'}X)^{+}X^{'}Y$ is noisy, where $(X^{'}X)^{+}$ is the Moore-Penrose generalized inverse of $X^{'}X$.

Vectorizing the matrices, we can transform model (1) into

$$y = (I_n \otimes X)\beta + e \tag{3}$$

where $y = vec(Y)$, $\beta = vec(B)$, $e = vec(E)$ and $e \sim N(0, \Sigma \otimes I_T)$. The dimension of β is $N^2k \times 1$.

We define the DAELasso estimator for a VAR as following:

$$\hat{\beta}_{dL} = arg\ min_{\beta}\{[y-(I_n \otimes X)\beta]'[y-(I_n \otimes X)\beta] + \sum_{j=1}^{N^2k} \lambda_{1,j}|\beta_j| + \sum_{j=1}^{N^2k} \lambda_{2,j}\beta_j^2\} \tag{4}$$

where $\lambda_{1,j}$ and $\lambda_{2,j}$, for $j = 1, 2, ..., N^2k$, are positive tuning parameters associated with the L_1 and L_2 penalties, respectively. We allow for different tuning parameters for different β_j to allow for different degree of shrinkages. For notational convenience, we define $\Lambda_1 = diag(\lambda_{1,1}, \lambda_{1,2}, ..., \lambda_{1,N^2k})$ and $\Lambda_2 = diag(\lambda_{2,1}, \lambda_{2,2}, ..., \lambda_{2,N^2k})$.

Traditional e-net Lasso imposes the same amount of shrinkage on all the coefficients, which tends to excessively penalize the coefficients for large effects (Zou, 2006; Zou and Zhang, 2009). In classical framework, adaptive Lasso and adaptive e-net Lasso are proposed to solve this problem (Zou, 2006; Zou and Zhang, 2009). Frequentist methods usually require first calculating different adaptive weights for different coefficients then attaching them to the final Lasso or e-net Lasso model to be estimated. This might cause problems if the weights derived in the first step are not the optimal candidates. By contrast, Bayesian adaptive Lasso methods such as those of Griffin and Brown (2010) and Leng et al (2010) can achieve adaptation and shrinkage in a single step. Griffin and Brown (2010) explore using different priors to achieve adaptive Lasso which automatically adapt. Leng et al (2010) propose using varying degree of shrinkage for the tuning parameter. Our adaptive approach is in spirit of Leng et al (2010).

2.1 Priors

Following the suggestions of Tibshirani (1996), univariate Bayesian Laplace prior, which can be expressed as a scale mixtures of Normals with an exponential density (Andrews and Mallows, 1974), is widely used to enforce sparsity induced by the L_1 penalty in Lasso (e.g, Park and Casella, 2008; Leng et al, 2010; Korobilis, 2011). It is natural to consider extending the univariate Bayesian Laplace prior into multivariate analysis. However, this is not so straightforward. As noted by van Gerven et al (2009, 2010), the commonly used multivariate Laplace distributions (e.g, Kotz et al, 2001; Eltoft el al, 2006) generally do not factorize into a product of univariate Laplace distributions that can be associated with the individual coefficients.

Our approach is directly motivated by van Gerven et al's (2009, 2010) multivariate Laplace prior for single equation models. van Gerven et al (2009, 2010) use a scale mixture of Normals to reflect their prior knowledge of the interactions between the coefficients. Our scale mixture prior is similar to theirs, however, our prior is more about ensuring the priors associated with the L_1 norm are conditional on the covariance matrix of the VAR errors (Σ). Conditioning on Σ is important because otherwise the posterior may not be unimodal (Park and Casella, 2008). The posterior of van Gerven et al (2009, 2010) is not in a tractable closed form, and they use approximate inference methods for posterior computations. By contrast, our prior can lead to closed-form posteriors that can be directly drawn from Gibbs sampler.

We consider a conditional multivariate mixture prior of the following form:

$$\pi(\beta|\Sigma,\Gamma,\Lambda_1,\Lambda_2) \propto \prod_{j=1}^{N^2k} \{\frac{\sqrt{\lambda_{2,j}}}{\sqrt{2\pi}} exp(-\frac{\lambda_{2,j}}{2}\beta_j^2) \\ \times \int_0^\infty \frac{1}{\sqrt{2\pi f_j(\Gamma))}} exp[-\frac{1}{2f_j(\Gamma)}\beta_j^2] d(f_j(\Gamma))\} \\ \times \{|M|^{-\frac{1}{2}} exp(-\frac{1}{2}\Gamma' M^{-1}\Gamma)\}^2 \tag{5}$$

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 $f_j(\Gamma)$. This is important as otherwise the posterior will be not unimodal due to the 'sharp corners' of the L_1 penalty (Park and and Casella, 2008). In (5), 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 (5) indicates that $\Gamma \sim N(0, M)$, with

$$M = \begin{pmatrix} M_{1,1} & ... & M_{1,j} & M_{1,j+1} & ... & M_{1,N^2k} \\ ... & ... & ... & ... & ... & ... \\ M_{j,1} & ... & M_{j,j} & M_{j,j+1} & ... & M_{j,N^2k} \\ M_{j+1,1} & ... & M_{j+1,j} & M_{j+1,j+1} & ... & M_{j+1,N^2k} \\ ... & ... & ... & ... & ... & ... \\ M_{N^2k,1} & ... & M_{N^2k,j} & M_{N^2k,j+1} & ... & M_{N^2k,N^2k} \end{pmatrix} \tag{6}$$

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).

$$\tau_1 = \gamma_1 + H_1(\gamma_2, \gamma_3, ..., \gamma_{N^2k})' \tag{7}$$

$$\tau_2 = \gamma_2 + H_2(\gamma_3, \gamma_4, ..., \gamma_{N^2k})' \tag{8}$$

...

$$\tau_{N^2K-1} = \gamma_{N^2k-1} + H_{N^2k-1}\gamma_{N^2k} \tag{9}$$

$$\tau_{N^2K} = \gamma_{N^2k} \tag{10}$$

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

$$N(\tau_1|0, \sigma^2_{\gamma_1})N(\tau_2|0, \sigma^2_{\gamma_2})...N(\tau_{N^2k}|0, \sigma^2_{\gamma_{N^2k}}) \tag{11}$$

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 (11) is 1. Defining $\eta_j = \tau_j/\lambda_{1,j}$, we can write (11) as

$$N(\eta_1|0, \sigma^2_{\gamma_1}\lambda^{-2}_{1,1})N(\eta_2|0, \sigma^2_{\gamma_2}\lambda^{-2}_{1,2})...N(\eta_{N^2k}|0, \sigma^2_{\gamma_{N^2k}}\lambda^{-2}_{1,N^2k}) \tag{12}$$

Let $f_j(\Gamma) = 2(\eta_j^2)$. Our scale mixture prior in (5) can be rewritten as:

$$\begin{aligned}\pi(\beta|\Sigma,\Gamma,\Lambda_1,\Lambda_2) \propto \prod_{j=1}^{N^2k} \{ & \frac{\sqrt{\lambda_{2,j}}}{\sqrt{2\pi}} exp(-\frac{\lambda_{2,j}}{2}\beta_j^2) \\ & \times \int_0^\infty \frac{1}{\sqrt{2\pi(2\eta_j^2)}} exp[-\frac{\beta_j^2}{2(2\eta_j^2)}] d(2\eta_j^2) \\ & \times \frac{\lambda_{1,j}^2}{2\sigma_{\gamma_j}^2} exp[-\frac{1}{2}\frac{2\eta_j^2}{(\sigma_{\gamma_j}^2)/\lambda_{1,j}^2}]\}\end{aligned} \tag{13}$$

The last two terms in (13) 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 (13) shows that the conditional prior for β_j is $N(0, \frac{2\eta_j^2}{2\lambda_{2,j}\eta_j^2+1})$, and the conditional prior for β is

$$\beta|\Gamma,\Sigma,\Lambda_1,\Lambda_2 \sim N(0, D_\Gamma^*) \tag{14}$$

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

Priors for Σ, $\lambda_{1,j}^2$ and $\lambda_{2,j}$ 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 $\lambda_{2,j}$: $\Sigma^{-1} \sim W(\underline{S}^{-1}, \underline{\nu})$, $\lambda_{1,j}^2 \sim G(\underline{\mu}_{\lambda_{1,j}^2}, \underline{\nu}_{\lambda_{1,j}^2})$, $\lambda_{2,j} \sim G(\underline{\mu}_{\lambda_{2,j}}, \underline{\nu}_{\lambda_{2,j}})$.[1]

[1] Please refer to Koop (2003), p326, for Gamma distribution, and Zellner (1971), p389, for Wishart distribution.

2.2 Posteriors and Gibbs Sampler

Combining the priors and likelihood, the following full conditional posteriors can be easily derived.

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 $\bar{\mu}_{\lambda_{1,j}} = \frac{\overline{\nu}_{\lambda_{1,j}} \sigma_j^2 \underline{\mu}_{\lambda_{1,j}}}{2\tau_j^2 \underline{\mu}_{\lambda_{1,j}} + \underline{\nu}_{\lambda_{1,j}} \sigma_j^2}$. The Full conditional posterior for $\lambda_{2,j}$ is $G(\bar{\mu}_{\lambda_{2,j}}, \bar{\nu}_{\lambda_{2,j}})$, where $\bar{\nu}_{\lambda_{2,j}} = \underline{\nu}_{\lambda_{2,j}} + 1$ and $\bar{\mu}_{\lambda_{2,j}} = \frac{\underline{\mu}_{\lambda_{2,j}} \bar{\nu}_{\lambda_{2,j}}}{\underline{\nu}_{\lambda_{2,j}} + \underline{\mu}_{\lambda_{2,j}} \beta_j^2}$. Finally the full conditional posterior of $\frac{1}{2\eta_j^2}$ is Inverse Gaussian: $IG(\sqrt{\frac{\lambda_{1,j}^2}{\beta_j^2 \sigma_{\gamma_j}^2}}, \frac{\lambda_{1,j}^2}{\sigma_{\gamma_j}^2})$. [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 $j = 1, 2, ...N^2k$
4. draw $\lambda_{2,j}|\beta, \Sigma, \Lambda_1, \Lambda_{2,-j}, \Gamma$ from $G(\bar{\mu}_{\lambda_{2,j}}, \bar{\nu}_{\lambda_{2,j}})$ for $j = 1, 2, ...N^2k$
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 $j = 1, 2, ...N^2k$.

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

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

3 Related Lasso Types of VAR Shrinkage

DAELasso provides a general method to shrink both the variable and parameter space of a VAR. However, with the number of tuning parameters two times the number of coefficients, DAELasso might be subject to the criticism of demanding too much from the data. In this section, we present four alternative scaled mixture priors for β that respectively associated with Lasso, adaptive Lasso, e-net Lasso, and adaptive e-net Lasso. Note that these four Lassos are all nested in DAELasso. For brevity, we are not to provide the posteriors as they can be easily worked out using the procedures presented for DAELasso shrinkage.

3.1 Lasso VAR Shrinkage

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

$$\hat{\beta}_L = arg\, min_\beta \{ [y - (I_n \otimes X)\beta]' [y - (I_n \otimes X)\beta] + \lambda_1 \sum_{j=1}^{N^2 k} |\beta_j| \} \tag{15}$$

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

$$\begin{aligned} \pi(\beta|\Sigma, \Gamma, \lambda_1) \propto \prod_{j=1}^{N^2 k} \{ \int_0^\infty \frac{1}{\sqrt{2\pi f_j(\Gamma))}} exp[-\frac{1}{2 f_j(\Gamma)} \beta_j^2] d(f_j(\Gamma)) \} \\ \times \{ |M|^{-\frac{1}{2}} exp(-\frac{1}{2} \Gamma' M^{-1} \Gamma) \}^2 \end{aligned} \tag{16}$$

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

$$\begin{aligned}\pi(\beta|\Sigma,\Gamma,\lambda_1) \propto \prod_{j=1}^{N^2k} \{ \int_0^\infty \frac{1}{\sqrt{2\pi(2\eta_j^2))}} exp[-\frac{\beta_j^2}{2(2\eta_j^2)}] d(2\eta_j^2) \\ \times \frac{\lambda_1^2}{2\sigma_{\gamma_j}^2} exp[-\frac{1}{2}\frac{2\eta_j^2}{(\sigma_{\gamma_j}^2)/\lambda_1^2}]\}\end{aligned} \tag{17}$$

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

3.2 Adaptive Lasso VAR Shrinkage

We define the adaptive Lasso estimator for a VAR as:

$$\hat{\beta}_{AL} = arg\, min_\beta \{ [y - (I_n \otimes X)\beta]' [y - (I_n \otimes X)\beta] + \sum_{j=1}^{N^2k} \lambda_{1,j} |\beta_j| \} \tag{18}$$

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

$$\begin{aligned}\pi(\beta|\Sigma,\Gamma,\Lambda_1) \propto \prod_{j=1}^{N^2k} \{ \int_0^\infty \frac{1}{\sqrt{2\pi f_j(\Gamma))}} exp[-\frac{1}{2f_j(\Gamma)}\beta_j^2] d(f_j(\Gamma)) \} \\ \times \{ |M|^{-\frac{1}{2}} exp(-\frac{1}{2}\Gamma' M^{-1}\Gamma) \}^2\end{aligned} \tag{19}$$

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

$$\begin{aligned}\pi(\beta|\Sigma,\Gamma,\Lambda_1) \propto \prod_{j=1}^{N^2k} \{ \int_0^\infty \frac{1}{\sqrt{2\pi(2\eta_j^2))}} exp[-\frac{\beta_j^2}{2(2\eta_j^2)}] d(2\eta_j^2) \\ \times \frac{\lambda_{1,j}^2}{2\sigma_{\gamma_j}^2} exp[-\frac{1}{2}\frac{2\eta_j^2}{(\sigma_{\gamma_j}^2)/\lambda_{1,j}^2}]\}\end{aligned} \tag{20}$$

where $\eta_j = \tau_j/\lambda_{1,j}$.

3.3 E-net Lasso VAR Shrinkage

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

$$\hat{\beta}_{EL} = arg\, min_{\beta}\{[y-(I_n \otimes X)\beta]'[y-(I_n \otimes X)\beta] + \lambda_1 \sum_{j=1}^{N^2k} |\beta_j| + \lambda_2 \sum_{j=1}^{N^2k} \beta_j^2\} \tag{21}$$

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

$$\begin{aligned} \pi(\beta|\Sigma,\Gamma,\lambda_1,\lambda_2) \propto \prod_{j=1}^{N^2k} \{ & \frac{\sqrt{\lambda_2}}{\sqrt{2\pi}} exp(-\frac{\lambda_2}{2}\beta_j^2) \\ & \times \int_0^{\infty} \frac{1}{\sqrt{2\pi f_j(\Gamma))}} exp[-\frac{1}{2f_j(\Gamma)}\beta_j^2] d(f_j(\Gamma))\} \\ & \times \{|M|^{-\frac{1}{2}} exp(-\frac{1}{2}\Gamma' M^{-1}\Gamma)\}^2 \end{aligned} \tag{22}$$

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

$$\begin{aligned} \pi(\beta|\Sigma,\Gamma,\lambda_1,\lambda_2) \propto \prod_{j=1}^{N^2k} \{ & \frac{\sqrt{\lambda_2}}{\sqrt{2\pi}} exp(-\frac{\lambda_2}{2}\beta_j^2) \\ & \times \int_0^{\infty} \frac{1}{\sqrt{2\pi(2\eta_j^2))}} exp[-\frac{\beta_j^2}{2(2\eta_j^2)}] d(2\eta_j^2) \\ & \times \frac{\lambda_1^2}{2\sigma_{\gamma_j}^2} exp[-\frac{1}{2}\frac{2\eta_j^2}{(\sigma_{\gamma_j}^2)/\lambda_1^2}]\} \end{aligned} \tag{23}$$

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

3.4 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:

$$\hat{\beta}_{AEL} = arg\, min_{\beta}\{[y-(I_n \otimes X)\beta]'[y-(I_n \otimes X)\beta] + \sum_{j=1}^{N^2k} \lambda_{1,j}|\beta_j| + \lambda_2 \sum_{j=1}^{N^2k} \beta_j^2\} \tag{24}$$

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

$$\begin{aligned} \pi(\beta|\Sigma,\Gamma,\Lambda_1,\lambda_2) \propto \prod_{j=1}^{N^2k} \{ & \frac{\sqrt{\lambda_2}}{\sqrt{2\pi}} exp(-\frac{\lambda_2}{2}\beta_j^2) \\ & \times \int_0^{\infty} \frac{1}{\sqrt{2\pi f_j(\Gamma))}} exp[-\frac{1}{2f_j(\Gamma)}\beta_j^2] d(f_j(\Gamma))\} \\ & \times \{|M|^{-\frac{1}{2}} exp(-\frac{1}{2}\Gamma' M^{-1}\Gamma)\}^2 \end{aligned} \tag{25}$$

Let $f_j(\Gamma) = 2(\eta_j^2)$. The scale mixture prior in (25) can be rewritten as:

$$\begin{aligned} \pi(\beta|\Sigma,\Gamma,\Lambda_1,\lambda_2) \propto \prod_{j=1}^{N^2k} \{ & \frac{\sqrt{\lambda_2}}{\sqrt{2\pi}} exp(-\frac{\lambda_2}{2}\beta_j^2) \\ & \times \int_0^{\infty} \frac{1}{\sqrt{2\pi(2\eta_j^2))}} exp[-\frac{\beta_j^2}{2(2\eta_j^2)}] d(2\eta_j^2) \\ & \times \frac{\lambda_{1,j}^2}{2\sigma_{\gamma_j}^2} exp[-\frac{1}{2}\frac{2\eta_j^2}{(\sigma_{\gamma_j}^2)/\lambda_{1,j}^2}]\} \end{aligned} \tag{26}$$

where $\eta_j = \tau_j/\lambda_{1,j}$.

4 Empirical Illustration

4.1 Data

In macroeconomics, it is a standard practice to assess models by their forecasting performance (e.g. Litterman, 1986; Giannone et al, 2010). Koop (2011) provides an extensive forecasts evaluation for seven popular Bayesian VAR priors. We employ the data set of Koop (2011), an updated version of that used in Stock and Watson (2008), for the out-of sample forecasting analysis. The data set contains twenty quarterly macroeconomic series including a measure of economic activity (GDP, real GDP), prices (CPI, the consumer price index), an interest rate (FFR, the Fed funds rate), and other seventeen variables.[3] Four of the seventeen variables are those used in the monetary model of Christiano el al (1999). The rest of the thirteen variables contain important aggregated information of the economy. The time series span from 1959Q1 to 2008Q4. A full list of the variables is provided in the Appendix. Detailed data descriptions please refer to Koop (2011) and Stock and Watson (2008). Data are transformed to stationarity and standardized same as Koop (2011).[4]

[3]These 20 variables are used for medium-size VAR in Koop (2011). Koop (2011) also examines the VAR forecasts using medium-large VAR, which contains 40 variables, and large VAR, which contains 168 variables. We only focus on Koop's (2011) medium-size VAR in this paper due to two considerations. First, it is computationally costly to use DAELasso priors to estimate the medium-large and large VARs. Second, it is shown in the literature (e.g. Banbura et al, 2010; Koop, 2011) that most of the gains in forecasting performance are achieved by using medium VARs of about 20 key variables.

[4]I am grateful to Mark Watson for providing the data. In addition, I am grateful to Gary Koop for sharing the Matlab code for data transformation.

4.2 Forecast Evaluation

Same as Koop (2011), we conduct rolling and recursive forecast exercises and calculate both the mean squared forecast error (MSFE) and predictive likelihood measures using reduced form VAR of order four. The window length for the rolling estimation is set to be ten years. Recursive and rolling forecasts are conducted for t_0+h, t_0+1+h,...T, where t_0 is 1969Q4. Let y^f_{t+h} be the h^{th} period forecast of y using data available at time t, and y_{t+h} be the real value for y observed at $t+h$. The MSFE measure for the variable y_i is calculated as an average of the mean squared errors of the point estimates:

$$MSFE = \frac{\sum_{t=t_0}^{T-h}[y_{i,t+h} - E(y^f_{i,t+h}|Data_t)]^2}{T-h-t_0+1} \tag{27}$$

The predictive likelihood is used to evaluate the entire predictive distribution. In particular, the following sum of the log predictive likelihood is used:

$$\sum_{t=t_0}^{T-h} log[p(y^f_{i,t+h} = y_{i,t+h}|Data_t)] \tag{28}$$

For DAELasso, we need to elicit priors for $\lambda^2_{1,j}$, $\lambda_{2,j}$, and Σ. It is practically impossible to set informative priors for each $\lambda^2_{1,j}$ and $\lambda_{2,j}$, thus we set relatively uninformative priors for $\lambda^2_{1,j}$ and $\lambda_{2,j}$ to be $G(1, 0.001)$ and $G(1, 0.01)$, respectively. The prior for Σ^{-1} is set to be $W(\frac{1}{N-1}I_N, 1)$, which is also relatively uninformative. There is room for improving the forecasting performance of DAELasso, in particular by eliciting more informative priors. We do not explore this possibility in the current paper because the goal of our exercise is to find out whether a DAELasso with relatively uninforma-

tive priors can provide acceptable forecasting results. For comparison, the priors for Lasso, adaptive Lasso, e-net Lasso, and adaptive e-net Lasso are set in the same manner.

Tables 1-4 report the DAELasso forecasts results along with Lasso, adaptive Lasso, e-net Lasso, adaptive e-net Lasso, and those of 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'. We refer to Koop (2011) for a lucid description of these priors.

Table 1: Rolling Forecasting for $h = 1$

	GDP	CPI	FFR
DAELasso	0.5774	0.3172	0.5730
	(-198.88)	(-192.66)	(-211.66)
adaptive e-net Lasso	0.6739	0.3976	0.6312
	(-195.78)	(-199.37)	(-215.03)
e-net Lasso	0.6751	0.3994	0.6326
	(-215.29)	(-211.6)	(-223.72)
adaptive Lasso	0.7663	0.3076	0.6184
	(-225.56)	(-209.24)	(-228.28)
Lasso	0.6697	0.3918	0.6263
	(-255.82)	(-241.26)	(-257.64)
Minn. Prior as in BGR	0.5842	-0.3414	0.5071
	(-190.51)	(-209.19)	(-177.41)
Minn. Prior Σ diagonal	0.6112	0.3048	0.5228
	(-194.04)	(-193.00)	(-181.74)
Minn. Prior Σ not diagonal	0.6092	0.3068	0.526
	(-192.12)	(-202.38)	(-185.88)
SSVS Conjugate semi-automatic	0.8061	0.3808	0.6324
	(-209.44)	(-231.81)	(-175.82)
SSVS Conjugate plus Minn. Prior	0.5916	0.3516	0.5069
	(-191.41)	(-212.11)	(-179.18)
SSVS Non-conj. semi-automatic	0.8780	0.4674	0.7348
	(-234.25)	(-235.98)	(-213.03)
SSVS Non-conj. plus Minn. Prior	0.6766	0.3402	0.5239
	(-197.85)	(-195.22)	(-177.18)

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	0.5545	0.4798	0.6458
	(-206.92)	(-205.93)	(-230.87)
adaptive e-net Lasso	0.5287	0.4745	0.5495
	(-195.68)	(-204.38)	(-219.92)
e-net Lasso	0.5288	0.4749	0.5494
	(-215.24)	(-213.47)	(-225.49)
adaptive Lasso	0.7428	0.5411	0.7776
	(-233.92)	(-223.04)	(-247.68)
Lasso	0.5286	0.4739	0.5511
	(-255.88)	(-242.64)	(-258.96)
Minn. Prior as in BGR	0.5852	0.5484	0.5847
	(-217.08)	(-227.69)	(-213.38)
Minn. Prior Σ diagonal	0.5869	0.5499	0.5882
	(-211.06)	(-232.37)	(-246.61)
Minn. Prior Σ not diagonal	0.578	0.5825	0.5825
	(-210.55)	(-222.21)	(-212.12)
SSVS Conjugate semi-automatic	1.2338	0.9892	1.3205
	(-282.63)	(-284.27)	(-273.83)
SSVS Conjugate plus Minn. Prior	0.6308	0.5371	0.6059
	(-230.23)	(-221.23)	(-213.5)
SSVS Non-conj. semi-automatic	1.5985	1.2194	1.638
	(-294.11)	(-266.18)	(-268.75)
SSVS Non-conj. plus Minn. Prior	0.6346	0.5097	0.583
	(-209.89)	(-201.3)	(-198.07)

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	0.5521	0.2882	0.5642
	(-210.43)	(-190.91)	(-224.24)
adaptive e-net Lasso	0.6739	0.3968	0.6313
	(-241.99)	(-201.55)	(-239.81)
e-net Lasso	0.675	0.3993	0.6325
	(-225.62)	(-212.49)	(-237.91)
adaptive Lasso	0.6248	0.275	0.5965
	(-219.17)	(-196.36)	(-226.84)
Lasso	0.6684	0.3886	0.6245
	(-236.48)	(-221.09)	(-242.93)
Minn. Prior as in BGR	0.5552	0.3029	0.5136
	(-192.29)	(-195.9)	(-229.14)
Minn. Prior Σ diagonal	0.5774	0.2756	0.5355
	(-204.84)	(-182.18)	(-238.79)
Minn. Prior Σ not diagonal	0.5489	0.2664	0.5164
	(-195.4)	(-184.06)	(-249.09)
SSVS Conjugate semi-automatic	0.6776	0.2724	0.6329
	(-199.9)	(-191.15)	(-245.25)
SSVS Conjugate plus Minn. Prior	0.5577	0.3088	0.5134
	(-192.53)	(-197.64)	(-228.54)
SSVS Non-conj. semi-automatic	0.6407	0.3161	0.579
	(-205.12)	(-196.47)	(-237.16)
SSVS Non-conj. plus Minn. Prior	0.6466	0.291	0.5431
	(-203.92)	(-187.58)	(-228.86)

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	0.5434	0.4789	0.6136
	(-218.34)	(-206.60)	(-239.62)
adaptive e-net Lasso	0.5288	0.4745	0.5495
	(-215.61)	(-207.02)	(-247.3)
e-net Lasso	0.5288	0.4749	0.5494
	(-225.54)	(-213.7)	(-238.98)
adaptive Lasso	0.6297	0.5165	0.664
	(-227.95)	(-214.7)	(-242.16)
Lasso	0.5288	0.4731	0.5496
	(-236.24)	(-222.75)	(-244.27)
Minn. Prior as in BGR	0.6094	0.5217	0.5868
	(-214.71)	(-219.35)	(-249.63)
Minn. Prior Σ diagonal	0.61	0.5191	0.6075
	(-214.02)	(-217.6)	(-278.11)
Minn. Prior Σ not diagonal	0.6214	0.5203	0.5882
	(-213.28)	(-216.07)	(-244.77)
SSVS Conjugate semi-automatic	0.6473	0.6042	0.5873
	(-212.35)	(-225.02)	(-249.46)
SSVS Conjugate plus Minn. Prior	0.8357	0.7031	0.6716
	(-219.64)	(-246.64)	(-258.47)
SSVS Non-conj. semi-automatic	0.7375	0.7723	0.8811
	(-293.21)	(-226.36)	(-268.06)
SSVS Non-conj. plus Minn. Prior	0.667	0.4883	0.5282
	(-219.01)	(-201.62)	(-233.67)

Notes:

MSFEs as proportion of random walk MSFEs.

Sum of log predictive likelihoods in parentheses.

Results presented in Tables 1-4 show that the forecasting performance of DAELasso approach is comparable to that of the seven popular Bayesian shrinkage methods explored in Koop (2011). Compared to the results reported in Koop (2011), DAELasso tends to forecast better for GDP and CPI in terms of point forecasts. In particular DAELasso provides the best rolling and recursive GDP and CPI point forecasts for $h = 4$. By contrast, DAELasso's point forecasts for FFR are not as good. However, the results are not too bad as DAELasso's performance generally ranks in the middle. The results for predictive loglikelihood yielded by DAELasso approach are more mixed. Yet, DAELasso's performances are qualitatively similar to that of the Bayesian VAR methods studied in Koop (2011).

When we focus on the five Lasso types of forecasts, we find for $h = 1$, DAELasso tends to provide the best GDP and FFR point forecasts while adaptive Lasso gives the best CPI point forecasts. This results seem to suggest that while GDP and FFR can be better forecasted by using many variables that might be highly collinear, CPI can be better forecasted using a smaller number of important variables that are not highly correlated with each other. Turning to the point forecasts for $h = 4$, we find more parsimonious models such as Lasso are performing better in most of the cases. This is in consistent with the general notion that parsimonious models are more valuable when the forecasting horizon gets longer. Forecasts results for predictive loglikelihood varies for the five Lasso types of methods. But overall, DAELasso and adaptive e-net Lasso tend to outperform the rest methods in most of the cases. This result suggest the importance of using e-net to capture the possible grouping effect, and using different degree of

shrinkages for different coefficients.

5 Conclusion

This paper proposes a Bayesian DAELasso approach for VAR shrinkage. We elicit a scale mixture prior which leads to closed-form posteriors that can be directly drawn from a Gibbs sampler. The method is appealing as it can simultaneously achieve variable selection and coefficient shrinkage in a data based fashion. DAELasso can constructively deal with multicollinearity problem by encouraging the grouping effect through e-net. In addition, it achieves adaption through allowing for different degree of shrinkages for different coefficients. Using relatively uninformative prior, we find that the forecasting performance of DAELasso is comparable to that of other popular Bayesian VAR shrinkage methods. This shows that DAELasso approach can be used as an appropriate addition to the available Bayesian VAR toolkits. The implementation of DAELasso approach is simple and straightforward. It can be easily extended into nonlinear framework to shed new light on macro economic analysis and forecasting.

References

[1] Anderson, T. W. (2003), An introduction to multivariate statistical analysis, *3rd Ed.* John Wiley and Sons: New Jersey.

[2] Andrews, D. F. and C. L. Mallows (1974), Scale mixtures of Normal distributions, *Journal of the Royal Statistical Society*, B 36, 99-102.

[3] Banbura, M, D. Giannone and L. Reichlin (2010), Large Bayesian vector auto regressions, *Journal of Applied Econometrics*, 25, 71-92.

[4] Burns, A. F and W. C. Mitchell (1946), Measuring Business Cycles, New York: NBER.

[5] Christiano, L. J., M. Eichenbaum and C. L. Evans (1999), Monetary policy shocks: what have we learned and to what end? in Taylor J. B. and M. Woodford (*eds*) Handbook of Macroeconomics 1, 65-148, Elsevier: Amsterdam.

[6] Chudik, A. and Pesaran, M. H. (2011), Infinite dimensional VARs and factor models, *Journal of Econometrics*, 163, 1, 4-22.

[7] De Mol, C., Giannone, D. and Reichlin, L. (2008), Forecasting using a large number of predictors: is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, 146, 318-28.

[8] Doan, T., R. Litterman and C. Sims (1984), Forecasting and conditional projections using realistic prior distributions, *Econometric Reviews*, 3, 1-100.

[9] Eltoft, T., T. Kim, and T. Lee (2006), On the multivariate Laplace distribution, *IEEE Signal Proc. Let.* 13 (5), 300-3.

[10] George, E. and R. McCulloch (1993), Variable selection via Gibbs sampling, *Journal of the American Statistical Association*, 88, 881-89.

[11] George, E. and R. McCulloch (1997), Approaches for Bayesian variable selection, *Statist. Sinica*, 7, 339-73.

[12] George, E., D. Sun and S. Ni (2008), Bayesian stochastic search for VAR model restrictions, *Journal of Econometrics*, 142, 553-80.

[13] Giannone, M., M. Lenza and G. Primiceri (2010), Prior selection for vector autoregressions, Working paper, Université Libre de Bruxelles.

[14] Griffin, J. E. and P. J. Brown (2010), Bayesian adaptive lassos with non-convex penalization, *Australian and New Zealand Journal of Statistics, forthcoming.*

[15] Kadiyala, K.R. and S. Karlsson (1997), Numerical methods for estimation and inference in Bayesian VAR-models, *Journal of Applied Econometrics*, 12(2), 99-132.

[16] Koop, G. (2003), Bayesian econometrics, John Wiley and Sons: Chichester.

[17] Koop, G. (2011), Forecasting with medium and large Bayesian VARs, *Journal of Applied Econometrics*, doi: 10.1002/jae.1270.

[18] Korobilis, D. (2011), Hierarchical shrinkage priors for dynamic regressions with many predictors, *MPRA* Paper No. 30380.

[19] Kotz, S., T. J. Kozubowski and K. Podgórski (2001), The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance, Birkhäuser Boston.

[20] Kwon, D.W., Landi, M.T., Vannucci, M., Issaq, H.J., Prieto, D. and Pfeiffer, R.M. (2011), An efficient stochastic search for Bayesian vari-

able selection with high-dimensional correlated predictors, *Computational Statistics and Data Analysis*, 55(10), 2807-18.

[21] Kyung, M., J. Gill, M. Ghosh, and G. Casella (2010), Penalized Regression, Standard Errors, and Bayesian Lassos, *Bayesian Analysis*, 5, 369-412.

[22] Leng, C, M. N. Tran and D. Nott (2010), Bayesian adaptive Lasso, *arXiv:1009.2300v1*

[23] Litterman, R.(1986), Forecasting with Bayesian vector autoregressions five years of experience, *Journal of Business and Economics Statistics*, 4, 25-38.

[24] Muirhead R. J. (1982), Aspects of Multivariate Statistical Theory, Wiley: New York.

[25] Park, T. and G. Casella (2008), The Bayesian Lasso, *Journal of the American Statistical Association*, 103, 681-86.

[26] Sims, C. (1972), Money, Income and Causality, *American Economic Review*, 62, 540-52.

[27] Sims, C. (1980), Macroeconomics and Reality, *Econometrica*, 48, 1-48.

[28] Song, S. and P. Bickel (2010), Large vector auto regressions, *arXiv:1106.3915v1.*

[29] Stock, J. H. and M. W. Watson (1989), New Indexes of Coincident and Leading Economic Indicators, *NBER Macroeconomic Annual* , 351-94.

[30] Stock, J. H. and M. W. Watson (2008), Forecasting in dynamic factor models subject to structural instability, in Castle, J. and N. Shephard (*eds*) The Methodology and Practice of Econometrics: A Festschrift in Honour of Professor David F. Hendry, 173205. Oxford University Press: Oxford.

[31] Tibshirani, R. (1996), Regression shrinkage and selection via the Lasso, *Journal of the Royal Statistical Society*, B 58, 267-88.

[32] van Gerven, M., B. Cseke, R. Oostenveld and T. Heskes (2009), Bayesian Source Localization with the Multivariate Laplace Prior, in Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams and A. Culotta *(eds.)* Advances in Neural Information Processing Systems 22, 1901-09.

[33] van Gerven, M., B. Cseke, F. P. de Lange and T. Heskes (2010), Efficient Bayesian multivariate fMRI analysis using a sparsifying spatio-temporal prior, *NeuroImage*, 50, 150-61.

[34] Yuan, M. and Y. Lin (2007), Model selection and estimation in regression with grouped variables, *Journal of the Royal Statistical Society*, B 68(1), 49-67.

[35] Zellner, A. (1971). An introduction to Bayesian inference in econometrics, John Wiley and Sons: New York.

[36] Zou, H. (2006), The adaptive Lasso and its oracle properties, *Journal of the American Statistical Association*, 101, 1418-29

[37] Zou, H. and T. Hastie (2005), Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society*, B 67, 301-20.

[38] Zou, H. and Zhang, H. H. (2009), On the adaptive elastic-Net with a diverging number of parameters, *Ann. Statist.*, 37, 1733-51.

Appendix: Data List

We use 20 US macro series used in Koop (2011), which is an updated version of Stock and Watson (2008). The variables are transformed to stationarity and standardized same as Koop (2011). Below we cite the essential data description and transformation code from Koop (2011). Please refer to Koop (2011) for detailed data information.

The data transformation codes are as following: 1: no transformation; 2: first difference; 3: second difference; 4: log; 5: first difference of logged variables; 6: second difference of logged variables.

Table 5: The List of Variables

Short name	Trans. Code	Data Description
RGDP	5	Real GDP, quantity index (2000 = 100)
CPI	6	CPI all items
FFR	2	Interest rate: federal funds (e-ective) (% per annum)
Com: spot price (real)	5	Real spot market price index: all commodities
Reserves nonbor	3	Depository inst reserves: nonborrowed (mil$)
Reserves tot	6	Depository inst reserves: total (mil$)
M2	6	Money stock: M2 (bil$)
Cons	5	Real Personal Cons. Exp., Quantity Index
IP: total	5	Industrial production index: total
Capacity Util	1	Capacity utilization: manufacturing (SIC)
U: all	2	Unemp. rate: All workers, 16 and over (%)
HStarts: Total	4	Housing starts: Total (thousands)
PPI: fin gds	6	Producer price index: finished goods
PCED	6	Personal Consumption Exp.: price index
Real AHE: goods	5	Real avg hrly earnings, non-farm prod. worker
M1	6	Money stock: M1 (bil$)
S&P: indust	5	S&Ps common stock price index: industrials
10 yr T-bond	2	Interest rate: US treasury const. mat., 10-yr
Ex rate: avg	5	US effective exchange rate: index number
Emp: total	5	Employees, nonfarm: total private