Comput Econ (2018) 51:1123–1138
https://doi.org/10.1007/s10614-017-9660-1

Sparse Bayesian Variable Selection in Probit Model for Forecasting U.S. Recessions Using a Large Set of Predictors

Yang Aijun[1,2] · Xiang Ju[3] · Yang Hongqiang[1] · Lin Jinguan[4]

Accepted: 20 January 2017 / Published online: 2 February 2017

Abstract In this paper, a large set of macroeconomic and financial predictors is used to forecast U.S. recession periods. We propose a sparse Bayesian variable selection in probit model for predicting U.S. recessions. The correlation prior is assigned for the binary vector to distinguish models with the same size, and the sparse prior is specified for the coefficient parameters for the purpose of predicting accurately using fewer parameters. In terms of the quadratic probability score and the log probability score, we demonstrate that the proposed method performs better than other three methods.

Keywords Sparse Bayesian variable selection · Correlation prior · Probit model · Forecasting U.S. recessions

1 Introduction

Recession forecasting is considered to be of special interest in a substantial amount of macroeconomic research. Knowing whether in the next month or next year the economy will be in an expansion or recession is an important piece of information for central bankers, entrepreneurs, and consumers (Fornaro 2016). For example, on the

✉ Yang Aijun
ajyang81@163.com

1 College of Economics and Management, Nanjing Forestry University, Nanjing, China

2 School of Economics and Management, Southeast University, Nanjing, China

3 Department of Financial Mathematics and Financial Engineering, South University of Science and Technology of China, Shenzhen, China

4 Institute of Statistics and Big Data, Nanjing Audit University, Nanjing, China

basis of what they predict will happen to the economy in the future, central banks can review their monetary policy and government authorities can tailor their spending.

In recent years, forecasting the values of the binary recession indicator with probit or logit models has attracted attention in the econometric literature. In these studies, only a few predictive variables at a time are used to forecast recession periods. For example, Estrella and Mishkin (1998) find that the spread between the 10-year Treasury bond rate and the three-month Treasury bill rate is the most useful predictor of future U.S. recessions. Wright (2006) finds that the level of the federal funds rate has some additional predictive power over and above the term spread. Katayama (2010) concludes that the combination of the term spread, month-to-month changes in the S&P 500 index, and the growth rate of non-farm employment gives better out-of-sample recession forecasting. However, the use of large panels for predicting recessions has not been widespread and recent literature has shown the value of large information for accurate economic forecasts (Stock and Watson 2002). Recently, Chen et al. (2011) and Fossati (2012) use the estimated latent factors extracted from a large number of time series to forecast future recessions. But these factor models based on large panels of macroeconomic indicators have some disadvantages. First, they require a two-step estimation procedure which has potential issues related to the generated regressor problem. Second, factors estimated from large panels have no clear economic interpretation. Including all potential predictors in a model is usually either infeasible or comes at a cost of great parameter estimation uncertainty, and also these predictors are highly correlated. Consequently methods which help to select the relevant predictors gain a lot of attention recently.

Bayesian methods offer a natural solution to overcome the curse of dimensionality problem by shrinking the parameters via the imposition of priors. In particular, Bayesian methods has been advocated as a device for identifying determinants of economic growth (Fernandez et al. 2001), exchange rates (Wright 2008), industrial production and consumer price index (De Mol et al. 2008), and inflation (Koop and Korobilis 2012). Bayesian variable selection methods with spike and slab priors have been proposed in the statistics literature (Mitchell and Beauchamp 1988; George and McCulloch 1993), and see George et al. (2008) and Korobilis (2012) for recent applications in linear macroeconomic forecasting models. These methods essentially differ in the form of the spike and slab priors. These methods have tried to achieve variable selection by indexing the predictors with a binary vector, and the number of ones in the binary vector (i.e. the model size) is binomially distributed. The prior for the binary vector can model the prior information on the model sizes, but do not distinguish models with the same size (Yuan and Lin 2005).

More recently, alternative sparse methods have been proposed for variable selection and forecasting. In the Bayesian framework, sparse methods are formulated by considering heavy-tailed priors for the regression coefficients. We can adjust the degree of sparseness by changing the prior distribution of the parameters. These methods assume that many parameters can be set to values very close to zero, and only a small subset of variables has an effect on the response. Such methods are preferable as they lead to more interpretable models and can predict more accurately with fewer variables. Many priors have been studied to eliminate the insignificant variables and so allow better selection of the significant variables. For example, the Laplace prior (Park and Casella 2008) (leading to Bayesian Lasso), and the Elastic-net prior (Li and Lin 2010) (leading to Bayesian Elastic Net). The Elastic-net is specifically designed to make sure to include all highly correlated variables. Fornaro (2016) adopt a Bayesian shrinkage approach in probit model for forecasting U.S. recession periods using a large set of macroeconomic and financial predictors. However, these methods inherit some problems: (a) Bayesian Lasso over-shrink large coefficients because of the relatively light tails of the priors. (b) The drawback of sparse Bayesian methods is that it perform only shrinkage of the regression coefficients towards to zero but do not automatically implement variable selection. (c) Highly correlated variables are needed to be avoided because those variables provide similar information for prediction (Yuan and Lin 2005).

There is a need for developing alternative sparse Bayesian variable selection method. In this paper, we propose an integrated sparse Bayesian variable selection in probit regression model for forecasting U.S. recessions. We perform variable selection through the Stochastic Search Variable Selection (SSVS; George and McCulloch 1993) technique. The generalized double Pareto prior (GDP; Armagan et al. 2013) is used for the coefficient parameters. The novelty of our method can be summarized as follows. Firstly, the GDP does not over-shrink coefficients that are not close to zero, and can be expressed as a scale mixture of normals that leads to a straightforward Gibbs sampler for posterior inferences. Secondly, the correlation prior (Yuan and Lin 2005) for the binary vector specified in this paper can distinguish models with the same size. Thirdly, it can automatically do the variable selection.

The remainder of the paper is organized as follows. In Sect. 2, we briefly review the probit model with latent variables, and discuss in details about the sparse prior and the implementation of the Bayesian method. Forecast evaluation is also discussed in this section. In Sect. 3 we illustrate the performance of our method on one real dataset. Section 4 provides a summary.

2 Method

2.1 Model

Throughout this study, we are interested in forecasting a binary variable $Y_{i+h}, i = 1, \ldots, n$, where $Y_{i+h} = 1$ indicates that the U.S. economy is in a recession during months $i + 1$ through $i + h$ and $Y_{i+h} = 0$ indicates that the U.S. economy is in an expansion during months $i+1$ through $i+h$. Let $X_i = (x_{i1}, \ldots, x_{ip})$ denotes a $(p \times 1)$ vector of the explanatory variables. We relate X_i with Y_i using a probit regression model i.e. $P(Y_{i+h} = 1) = \Phi(\alpha + X_i\beta)$, where α is the intercept, $\beta = (\beta_1, \ldots, \beta_p)^T$ is the vector of unknown regression parameters, Φ is the standard normal cumulative distribution function, and h is the forecasting horizon.

Following Albert and Chib (1993), we introduce the latent variables $Z = (Z_1, \ldots, Z_n)^T$ to convert the probit model to a linear regression model with inequality constraints on the latent variables. In the next stage, the latent variables Z_i are modelled as

$$Z_i = \alpha + X_i\beta + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{1}$$

where ε_i are independent and identically distributed $N(0, 1)$. The relationship between Y_i and Z_i is defined as

$$Y_{i+h} = \begin{cases} 1 & \text{if } Z_i > 0, \\ 0 & \text{if } Z_i \leq 0. \end{cases} \tag{2}$$

In order to perform the variable selection, a binary vector $\gamma = (\gamma_1, \ldots, \gamma_p)$ of p indicator variables is introduced and defined as

$$\gamma_i = \begin{cases} 1 & \text{if the } i\text{-th variable is included in the model,} \\ 0 & \text{if the } i\text{-th variable is excluded from the model.} \end{cases} \tag{3}$$

Given γ, let $p_\gamma = \sum_{i=1}^{p} \gamma_i$, β_γ be a p_γ by 1 vector consisting of all the nonzero elements of β, and $\mathbf{X}_\gamma$ be an n by p_γ matrix of covaraites consisting of all the columns of $\mathbf{X}$ corresponding to those elements of γ that are equal to 1. Adopting these notations, model (1) can be rewritten as

$$Z_i = \alpha + \mathbf{X}_{i,\gamma}\beta_\gamma + \varepsilon_i, \tag{4}$$

where $\mathbf{X}_{i,\gamma}$ is the i-th row of $\mathbf{X}_\gamma$.

2.2 Prior specification

The choice of the prior distributions for the unknown parameters is very important in the Bayesian method. In this paper, prior distributions for α, β_γ, and γ with the structure $p(\alpha, \beta_\gamma, \gamma) = p(\alpha)p(\beta_\gamma|\gamma)p(\gamma)$ is considered.

The prior distribution of α is taken as

$$\alpha \sim N(0, \mathbf{h}_\alpha), \tag{5}$$

where $\mathbf{h}_\alpha$ is a hyperparameter representing the variance of the univariate normal distribution. Since α is not our focus, a specified value is assigned to $\mathbf{h}_\alpha$. According to Lamnisos et al. (2009), a large value of $\mathbf{h}_\alpha$ is taken.

Now we specify the priors for the regression coefficients. By the definition of γ_i, it is natural to force $\beta_i = 0$ if $\gamma_i = 0$. But if $\gamma_i = 1$, the GDP prior is used. Compared with other priors such as Student t and Laplace, this sparse prior has some appealing properties: (a) it has a spike at zero like the Laplace density and has a Student t-like tail behavior, thus it can lead to sparse point estimates and does not over-shrink coefficients that are not close to zero. (b) it has a simple analytic form and can be written as a scale mixture of normals, thus can yield a proper posterior and will facilitate straightforward Bayesian computation even in large p cases. The prior distribution for β_i given γ_i can be presented as a three-level hierarchical model. At the first level, the regression coefficient β_i given $\gamma_i = 1$ is assumed to follow $\beta_i|\gamma_i = 1 \sim N(0, \lambda_i)$. At the second level, an exponential prior is assumed for λ_i, $\lambda_i \sim \exp(\tau_i^2/2)$. At the third level, τ_i is assumed to follow Gamma(a, b) distribution, where a, b are hyperparameters.

Let $|\gamma| = \sum_{i=1}^{p} \gamma_i$. For γ, a widely used prior is $p(\gamma) = \theta^{|\gamma|}(1-\theta)^{p-|\gamma|}$, which assumes that each variable is included in the model independently with probability θ, whether or not the variables are correlated. This prior can model the prior information on the model sizes, but can not distinguish models with the same size. Also since highly correlated variables provide similar information on the response, it is often the case that those variables are to be avoided. In this respect, we consider the following correlation prior for γ

$$p(\gamma) \propto \theta^{|\gamma|}(1-\theta)^{p-|\gamma|}\sqrt{\det(\mathbf{X}_\gamma^T\mathbf{X}_\gamma)}, \tag{6}$$

where $\det(\mathbf{X}_\gamma^T\mathbf{X}_\gamma) = 1$ if $|\gamma| = 0$. With respect to the effect of correlation between covariates in the prior specification, we consider the following conditional prior odds ratio for $\gamma_i = 1$

$$\frac{p(\gamma_i = 1|\gamma_{(-i)})}{p(\gamma_i = 0|\gamma_{(-i)})} = \frac{\theta}{1-\theta}\sqrt{\frac{\det(\mathbf{X}^T_{\gamma_i=1,\gamma_{(-i)}}\mathbf{X}_{\gamma_i=1,\gamma_{(-i)}})}{\det(\mathbf{X}^T_{\gamma_i=0,\gamma_{(-i)}}\mathbf{X}_{\gamma_i=0,\gamma_{(-i)}})}}, \tag{7}$$

where $\gamma_{(-i)}$ denotes the vector of γ with the i-th element deleted. We can see that the second factor of the right side of above equation will be small when X_i is strongly correlated with $\mathbf{X}_{\gamma_i=0,\gamma_{(-i)}}$, and consequently X_i can effectively be excluded from the whole model, because X_i does not contain much additional information.

2.3 Computation

Denote $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_p)$, $\tau = (\tau_1, \ldots, \tau_p)$. The full joint posterior distribution is given by

$$\begin{aligned} p\left(Z, \alpha, \beta_\gamma, \Lambda, \tau, \gamma|Y, \mathbf{X}\right) &\propto \exp\left\{-\frac{\sum_{i=1}^{n}\left(Z_i - \alpha - X_{i,\gamma}\beta_\gamma\right)^2}{2}\right\}\prod_{i=1}^{n}\mathrm{I}\left(A_i\right) \\ &\times \exp\left(-\frac{\alpha^2}{2h_\alpha}\right) \times \prod_{i\in m(\gamma)}\lambda_i^{-\frac{1}{2}}\exp\left(-\frac{\beta_i^2}{2\lambda_i}\right) \times \prod_{i=1}^{p}\tau_i^2\exp\left(-\frac{\tau_i^2\lambda_i}{2}\right) \\ &\times \prod_{i=1}^{p}\tau_i^{a-1}\exp\left(-\tau_i b\right) \times \theta^{|\gamma|}\left(1-\theta\right)^{p-|\gamma|}\sqrt{\det\left(\mathbf{X}_\gamma^T\mathbf{X}_\gamma\right)}, \end{aligned} \tag{8}$$

where A_i is equal to either $\{Z_i : Z_i > 0\}$ or $\{Z_i : Z_i \leq 0\}$ corresponding to $Y_i = 1$ or $Y_i = 0$, respectively; $\mathrm{I}(\cdot)$ is an indicator function; and $m(\gamma)$ is the subscript set of those elements of γ that are equal to 1 are included.

The posterior distribution is not available in explicit form and cannot be analytically evaluated, the MCMC method (Gilks et al. 1996), specifically Gibbs sampling (Geman and Geman 1984) is used to generate parameters from the conditional distributions. Since γ encapsulates the effectiveness of the different variables in explaining the variation in the responses, the posterior distribution of γ is of particular interest. Because α is rarely of interest, we marginalize it out for the purpose of simplicity and speed (Park and Casella 2008). The number of possible variables p is large, and in order to make the sampling scheme eciently explore the space of 2^p variables, we jointly update correlated components to explore the posterior distributions and search for high probability γ values. $Z, \beta_\gamma, \Lambda, \tau$ and γ are updated in turn based on $p(Z, \Lambda|\mathbf{X}, Y, \beta, \tau, \gamma) \propto p(Z|\mathbf{X}, Y, \Lambda, \gamma)p(\Lambda|\beta, \tau, \gamma)$ and $p(\beta_\gamma, \tau, \gamma|\mathbf{X}, Z, \Lambda) \propto p(\beta_\gamma|\mathbf{X}, Z, \Lambda, \gamma)p(\gamma|\mathbf{X}, Z, \Lambda)p(\tau|\beta)$.

The conditional distributions for implementing our sampling scheme are given below:

(i) The conditional posterior distribution of Z, $p(Z|\mathbf{X}, Y, \Lambda, \gamma)$, is given by

$$p(Z|\mathbf{X}, Y, \Lambda, \gamma) \propto N(0, \Sigma_\gamma)\prod_{i=1}^{n} \mathrm{I}(A_i), \tag{9}$$

with $\Sigma_\gamma = h_\alpha 1_n 1_n^T + \mathbf{X}_\gamma \Lambda_\gamma \mathbf{X}_\gamma^T + \mathbf{I}_n$, which is a multivariate truncated normal distribution. In (9), β is marginalized out from the posterior distribution $p(Z|\mathbf{X}, Y, \beta, \Lambda, \gamma)$ to reduce autocorrelation between β and Z, thus to improve mixing in the Markov chain. Direct sampling from (9) is known to be difficult. We follow the method of Devroye (1986) to simulate samples from the univariate truncated normal distribution $p(Z_i|Z_{(-i)}, \mathbf{X}, Y, \Lambda, \gamma)$, where $Z_{(-i)}$ is the vector of Z without the i-th element.

(ii) The posterior distribution of the i-th diagonal element of Λ is

$$\lambda_i^{-1}|\beta_i, \gamma_i, \tau_i \sim \mathrm{InvGauss}\left(\frac{\sqrt{\tau_i^2}}{|\beta_i|}, \tau_i^2\right), \tag{10}$$

where InvGauss denotes the inverse Gaussian distribution with the probability density function

$$\mathrm{InvGauss}(\iota, \kappa) = \sqrt{\frac{\kappa}{2\pi u^3}}\exp\left\{-\frac{\kappa(u-\iota)^2}{2\iota^2 u}\right\}, u > 0. \tag{11}$$

We use the algorithm given in Chhikara and Folks (1989) to generate the random observations from the inverse Gaussian distribution.

(iii) The full conditional distribution of β_γ is as follows

$$\beta_\gamma|\mathbf{X}, Z, \Lambda, \gamma \sim N(\Omega_\gamma \mathbf{X}_\gamma^T \Phi Z, \Omega_\gamma), \tag{12}$$

where $\Phi = (h_\alpha 1_n 1_n^T + \mathbf{I}_n)^{-1}$, and $\Omega_\gamma = (\mathbf{X}_\gamma^T \Phi \mathbf{X}_\gamma + \Lambda_\gamma^{-1})^{-1} = \Lambda_\gamma - \Lambda_\gamma \mathbf{X}_\gamma^T \Phi(\Phi \mathbf{X}_\gamma \Lambda_\gamma \mathbf{X}_\gamma^T \Phi + \Phi)^{-1} \Phi \mathbf{X}_\gamma \Lambda_\gamma$. The Woodbury-Sherman-Morrison matrix identity is used to reduce the dimension of the matrix, from p to n, and can make the computation much faster when data are high-dimensional with small sample size.

(iv) The conditional distribution of $\tau_i|\beta_i, \gamma$ is $\mathrm{Ga}(a+1, |\beta_i|+b)$.

(v) The conditional distribution of γ, $p(\gamma|\mathbf{X}, Z, \Lambda)$, is proportional to $|\Sigma_\gamma|^{-\frac{1}{2}}\exp(-\frac{Z^T\Sigma_\gamma^{-1}Z}{2}) \times \theta^{|\gamma|}(1-\theta)^{p-|\gamma|}\sqrt{\det(\mathbf{X}_\gamma^T\mathbf{X}_\gamma)}$. According to Panagiotelisa and Smith (2008), β is marginalized out from the conditional distribution $p(\gamma|\mathbf{X}, Z, \beta, \Lambda)$ so that the Markov chain would be non-reducible. To build an efficient Gibbs sampler, instead of drawing γ as a vector, it is better to draw a component γ_i of γ conditionally on $\gamma_{(-i)}$, where $\gamma_{(-i)}$ is the vector of γ without the i-th element, and

$$p\left(\gamma_i|\gamma_{(-i)}, \mathbf{X}, Z, \Lambda\right) \propto |\Sigma_\gamma|^{-\frac{1}{2}}\exp\left(-\frac{Z^T\Sigma_\gamma^{-1}Z}{2}\right)\theta^{\gamma_i}(1-\theta)^{1-\gamma_i}\sqrt{\det(\mathbf{X}_\gamma^T\mathbf{X}_\gamma)}. \tag{13}$$

Because γ_i only takes 0 or 1, we can consider the conditional probabilities of $p(\gamma_i = 1|\gamma_{(-i)}, \mathbf{X}, Z, \Lambda)$ and $p(\gamma_i = 0|\gamma_{(-i)}, \mathbf{X}, Z, \Lambda)$. Let $\gamma^1 = (\gamma_i, \ldots, \gamma_{i-1}, \gamma_i = 1, \gamma_{i+1}, \ldots, \gamma_p)$ and $\gamma^0 = (\gamma_i, \ldots, \gamma_{i-1}, \gamma_i = 0, \gamma_{i+1}, \ldots, \gamma_p)$, and similarly define Σ_{γ^1} and Σ_{γ^0} as Σ_γ in (9). After straightforward computation of (13), we have

$$p\left(\gamma_i = 1|\gamma_{(-i)}, \mathbf{X}, Z, \Lambda\right) = \left(1 + \frac{1-\theta}{\theta}\rho\right)^{-1}, \tag{14}$$

where

$$\rho = |\Sigma_{\gamma^1}\Sigma_{\gamma^0}^{-1}|^{\frac{1}{2}}\exp\left\{\frac{Z^T\left(\Sigma_{\gamma^1}^{-1} - \Sigma_{\gamma^0}^{-1}\right)Z}{2}\right\}\sqrt{\frac{\det\left(\mathbf{X}_{\gamma^0}^T\mathbf{X}_{\gamma^0}\right)}{\det\left(\mathbf{X}_{\gamma^1}^T\mathbf{X}_{\gamma^1}\right)}}. \tag{15}$$

We estimate the posterior probability of inclusion of each variable by calculating the number of times that each variable selected in the model through out the MCMC chain after the initial burn in.

$$\hat{p}(\gamma_i = 1|\mathbf{X}, Y) = \frac{1}{M}\sum_{k=1}^{M}\gamma_i^{(k)}, \tag{16}$$

variables with posterior inclusion probabilities larger than 0.5 are expected to be responsible for forecasting U.S. recessions.

2.4 Forecast evaluation

Following Wright (2006), let Y_{i+h} be a binary variable that takes on a value 1 if and only if there is an NBER-defined recession[1] at some point during months $i+1$ through

[1] The NBER defines a recession as a significant decline in economic activity spread across the economy, lasting more than a few months, normally visible in production, employment, real income, and other indicators.

$i + h$. A forecast of the probability of a recession in the next h months from a probit regression is then given by

$$\hat{p}_{i+h} = \int p(Y_{i+h} = 1|Y, \mathbf{X}, \Theta) p(\Theta|Y, \mathbf{X}) d\Theta, \tag{17}$$

where $\Theta = (Z, \Lambda, \beta_\gamma, \gamma, \tau)$ is the vector of some parameters. The above integral is intractable and we approximate it through Monte Carlo integration

$$\hat{p}_{i+h} = \frac{1}{M} \sum_{k=1}^{M} p(Y_{i+h} = 1|Y, \mathbf{X}, \Theta^{(k)}), \tag{18}$$

where $\Theta^{(k)}, k = 1, \ldots, M$, is the MCMC posterior sample of the parameter Θ after the initial burn in.

Out-of-sample forecasting results are evaluated using two statistics: the quadratic probability score (QPS) and the log probability score (LPS). QPS is equivalent to the mean squared error in the models for real-valued variables (Christiansen et al. 2014), and is defined as

$$QPS = \frac{2}{n^*} \sum_{t=1}^{n^*} (Y_{i+h} - \hat{p}_{i+h}), \tag{19}$$

where n^* is the effective number of out-of-sample forecasts. The value of the QPS is between 0 and 2, and smaller values indicate more accurate predictions. LPS is given by

$$LPS = -\frac{1}{n^*} \sum_{t=1}^{n^*} [Y_{i+h} log \hat{p}_{i+h} - (1 - Y_{i+h}) log(1 - \hat{p}_{i+h})], \tag{20}$$

the LPS can take values from 0 to ∞ and smaller values indicate more accurate predictions.

3 Empirical study

In this section, the effectiveness of our proposed Bayesian method, denoted as SBVS-PM, will be demonstrated with the help of empirical study. The performance of our Bayesian method will be compared with the following four methods: (A) the Bayesian probit model with the GDP prior on β, but no indicator variable γ; (B) the Bayesian probit model with the correlation prior for γ, but without additional shrinkage through the GDP prior (instead a standard normal prior is applied to β); (C) the Bayesian probit model with the GDP prior, but without the correlation term $\sqrt{\det(\mathbf{X}^T \mathbf{X})}$ in the prior for γ; (D) the method of Fornaro (2016). The reason to focus on models (A), (B) and (C) is that they represent different sparsity and correlation structure of data availability. Our proposed model can apply shrinkage twice and can take the correlation structure in **X** into account; model (A) can apply shrinkage once but cannot take the correlation structure in **X** into account; model (B) can apply shrinkage once and can take the correlation structure in **X** into account; model (C) can apply shrinkage twice but cannot take the correlation structure in **X** into account. In recent years, similar ideas of these three models have been successfully used for macroeconomic forecasting with a large number of highly correlated predictors; for recent work, see Korobilis (2013a, b), Gefang (2014), Belmonte et al. (2014), Stankiewicz (2015) and Korobilis (2016).

We compute recession forecasts using a monthly U.S. data, which starts in 1959:02 and ends in 2009:02. The data is divided into the in-sample part from 1959:02 to 1979:11 and out-of-sample part from 1979:12 to 2009:02. The predictive variables are taken from Stock and Watson (2012) dataset, which includes 108 variables. The reason of choosing this dataset is that many researchers have collected larger or smaller datasets but the coverage of the data is quite similar to the Stock-Watson data. The data includes 12 main categories: IP, employment, unemployment rate, housing, inventories, prices, wages, interest rates, money, exchange rates, stock prices and consumer expectations. Variables are transformed to achieve stationarity and standardized to have mean zero and standard deviation one. The full list of all variables and the related transformations is given in the "Appendix". We compute forecasts using an expanding window approach where the estimation window increases by one observation at each time when computing new forecasts. The results are shown for three h-step ahead forecast horizons: $h = 3$, 6, and 12 months. For the prior hyperparameters, we set $a = 1, b = 1, \theta = 0.5, h_\alpha = 100$. After monitoring for convergence, the Gibbs sampler is run for 150,000 iterations after an initial burn-in period of 50,000 iterations.

3.1 Empirical results

3.1.1 In-sample results

Even though we are mainly interested in out-of-sample forecasts it is useful to first compare the in-sample performance of different models over the entire sample period. To evaluate the performance of each model, the in-sample fit is evaluated using McFadden's pseudo-R2 (R^2_{mf}). The R^2_{mf} is defined as $R^2_{mf} = 1 - \frac{ln\hat{L}}{lnL_0}$, where $ln\hat{L}$ is the value of the log likelihood function evaluated at the estimated parameters, and lnL_0 is the value of the log likelihood function when all regression coefficients, except the intercept term, are zero. We also use the Bayesian Information Criterion (BIC) to decide the specification. Table 1 reports the in-sample R^2_{mf} and BIC for $h = 3, 6, 9$ months. The main results for $h = 3, 6, 9$ months are similar. Firstly, model (A) yields better in-sample fit than model (B). Secondly, model (B) gives better fit than model (C). Overall, the criteria already indicate that as expected, larger models achieve more accurate in-sample fit. For example, model (A) which imposes less shrinkage than model (C) gives the better in-sample fit. However, in-sample and out-of-sample fits are negatively correlated, which implies that good in-sample results do not guarantee good out-of-sample predictability. This over-fitting problem is particularly important, it is likely that models with high predictive accuracy in-sample may have poor forecast-

Table 1 In-sample R^2_{mf} and BIC of model (A), model (B), model (C), model (D) and SBVS-PM; the forecasting horizons $h = 3$, 6 and 12

	R^2_{mf}			BIC		
	h = 3	h = 6	h = 9	h = 3	h = 6	h = 9
Model (A)	0.48	0.51	0.52	−2.45	−2.40	−2.31
Model (B)	0.47	0.49	0.51	−2.44	−2.37	−2.30
Model (C)	0.46	0.48	0.50	−2.37	−2.32	−2.29
Model (D)	0.41	0.45	0.48	−2.34	−2.30	−2.26
SBVS-PM	0.39	0.42	0.45	−2.29	−2.25	−2.18

ing performance. Therefore, we conduct further out-of-sample prediction exercises in the next section.

3.1.2 Out-of-sample results

Table 2 reports the out-of-sample QPS and LPS for $h = 3$ months. These forecast evaluation statistics suggest that the out-of-sample performance of our proposed method is better than other three methods. Relative to model (C) which is the second best method, SBVS-PM reduces the QPS by 18% and the LPS by 12%. Tables 2 also reports the out-of-sample QPS and LPS for $h = 6$ and $h = 12$ months, respectively. The main results are similar to those found for $h = 3$. Firstly, our proposed method gives better out-of-sample fit than model (C), this might suggest that taking into account the correlation among the predictors is important with these data. Secondly our proposed method also gives better out-of-sample fit than model (B), thus improving sparsity does help to obtain better forecasting performance. The results from this out-of-sample forecasting exercise overall confirms the conclusion that our proposed method is the best fitting model at all horizons. This also confirms the need of sparsity and taking the correlation between the predictors into account when increasing the set of explanatory variables. This is expected as the models with a large number of predictors suffer from overfitting. Nowadays, large datasets are available to central banks, statistical offices and many other institutions, so being able to select the relevant information available to forecast the future state of the economy is highly beneficial. Sparse Bayesian exam-

Table 2 Out-of-sample QPS and LPS of model (A), model (B), model (C), model (D) and SBVS-PM; the forecasting horizons h=3, 6 and 12

	QPS			LPS		
	h = 3	h = 6	h = 9	h = 3	h = 6	h = 9
Model (A)	0.25	0.29	0.34	0.31	0.36	0.41
Model (B)	0.24	0.26	0.28	0.27	0.30	0.36
Model (C)	0.17	0.19	0.23	0.26	0.28	0.31
Model (D)	0.16	0.19	0.22	0.25	0.26	0.30
SBVS-PM	0.14	0.17	0.19	0.23	0.24	0.27

Table 3 Out-of-sample QPS and LPS of model (A), model (B), model (C) and SBVS-PM; the forecasting horizons $h = 3$, 6 and 12; prior setting (a) $\theta = 0.4$, $a = 1$, $b = 1$; prior setting (b) $\theta = 0.5$, $a = 2$, $b = 1$

		QPS			LPS		
		h = 3	h = 6	h = 9	h = 3	h = 6	h = 9
Prior setting (a)	Model (A)	0.26	0.29	0.34	0.31	0.36	0.41
	Model (B)	0.25	0.29	0.31	0.28	0.30	0.37
	Model (C)	0.18	0.20	0.23	0.26	0.29	0.32
	SBVS-PM	0.14	0.18	0.20	0.24	0.26	0.28
Prior setting (b)	Model (A)	0.27	0.31	0.36	0.33	0.39	0.43
	Model (B)	0.24	0.26	0.28	0.27	0.30	0.36
	Model (C)	0.18	0.21	0.24	0.26	0.29	0.33
	SBVS-PM	0.14	0.18	0.21	0.24	0.25	0.28

ined in this paper allows us to deal with large information set without incurring into the problem of overfitting and, as we have seen above, giving competitive out-of-sample forecasts.

The second exercise examines the robustness of the results obtained above for a few other specifications of the hyperparameters. Decreasing θ or b value or increasing a value would impose higher degree of sparsity of the models, as a result less predictors would be included in the models. To assess the sensitivity of the Bayesian results to the inputs of hyperparameters in the prior distributions, the empirical study was reanalyzed for: Prior setting (a) $\theta = 0.4$, $a = 1$, $b = 1$; Prior setting (b)$\theta = 0.5$, $a = 2$, $b = 1$. From Table 3 we can see that when the degree of sparsity becomes large, the differences between the four Bayesian methods for the long sample become quite small, and the results for the four Bayesian methods for the short sample are still similar as before. SBVS-PM still shows the best forecasting performance, model (C) tends to be better than model (A) and model (B). Therefore, it seems that the results of the empirical exercise are robust to the specifications of the hyperparameters when the degree of sparsity is moderate. Only when one can assume that the predictors have very little predictive power, imposing very high degree of sparsity on the coefficients will be reasonable. Otherwise, a moderate degree of sparsity seems more reasonable. Thus, the empirical analysis in this paper is carried out for moderate degree of sparsity.

4 Summary

The main focus of this paper is forecasting U.S. recessions in probit model with a large set of predictors. In this paper, we incorporate a Bayesian variable selection technique with the help of a binary vector for adaptive predictor selection. The correlation prior is assigned for the binary vector to distinguish models with the same size, and the sparse prior is specified for the coefficient parameters for the purpose of predicting accurately using fewer parameters. By representing the sparse prior as a three level hierarchical model, an effective MCMC algorithm is developed to generate samples from the posterior distributions. Based on the real dataset, we demonstrate that the proposed method performs better than other three methods in terms of the quadratic probability score and the log probability score.

Acknowledgements The authors gratefully acknowledge the financial support of the Natural Science Foundation of China (11501294, 11571073), the China Postdoctoral Science Foundation (2015M580374, 2016T90398), the Natural Science Foundation of Jiangsu (BK20141326) and the Research Fund for the Doctoral Program of Higher Education of China (20120092110021).

Appendix: Data description

The data used in Sect. 3 are presented here. The format is as follows: name, transformation code, and brief series description. The transformation codes are $1 = \text{notransformation}, 2 = \text{firstdifference}, 4 = \text{logarithm}, 5 = \text{firstdifferenceoflogarithms}, 6 = \text{seconddifferenceoflogarithms}$ (Table 4).

Table 4 Data descriptions

Name	Tran	Description
IPS10	5	INDUSTRIAL PRODUCTION INDEX-TOTAL INDEX
IPS11	5	INDUSTRIAL PRODUCTION INDEX-PRODUCTS, TOTAL
IPS299	5	INDUSTRIAL PRODUCTION INDEX-FINAL PRODUCTS
IPS12	5	INDUSTRIAL PRODUCTION INDEX-CONSUMER GOODS
IPS13	5	INDUSTRIAL PRODUCTION INDEX-DURABLE CONSUMER GOODS
IPS18	5	INDUSTRIAL PRODUCTION INDEX-NONDURABLE CONSUMER GOODS
IPS25	5	INDUSTRIAL PRODUCTION INDEX-BUSINESS EQUIPMENT
IPS32	5	INDUSTRIAL PRODUCTION INDEX-MATERIALS
IPS34	5	INDUSTRIAL PRODUCTION INDEX-DURABLE GOODS MATERIALS
IPS38	5	INDUSTRIAL PRODUCTION INDEX-NONDURABLE GOODS MATERIALS
IPS43	5	INDUSTRIAL PRODUCTION INDEX-MANUFACTURING (SIC)
IPS307	5	INDUSTRIAL PRODUCTION INDEX-RESIDENTIAL UTILITIES
IPS306	5	INDUSTRIAL PRODUCTION INDEX-FUELS
PMP	1	NAPM PRODUCTION INDEX (PERCENT)
UTL11	1	CAPACITY UTILIZATION-MANUFACTURING (SIC)
CES275	6	AVG HRLY EARNINGS, PROD WRKRS, NONFARM-GOODS-PRODUCING
CES277	6	AVG HRLY EARNINGS, PROD WRKRS, NONFARM-CONSTRUCTION
CES278	6	AVG HRLY EARNINGS, PROD WRKRS, NONFARM-MFG
CES275R	5	REAL AVG HRLY EARNINGS, PROD WRKRS, NONFARM-GOODS-PRODUCING (CES275/PI071)
CES277R	5	REAL AVG HRLY EARNINGS, PROD WRKRS, NONFARM-CONSTRUCTION (CES277/PI071)
CES278R	5	REAL AVG HRLY EARNINGS, PROD WRKRS, NONFARM-MFG(CES278/PI071)
CES002	5	EMPLOYEES, NONFARM-TOTAL PRIVATE

Table 4 continued

Name	Tran	Description
CES003	5	EMPLOYEES, NONFARM-GOODS-PRODUCING
CES006	5	EMPLOYEES, NONFARM-MINING
CES011	5	EMPLOYEES, NONFARM-CONSTRUCTION
CES015	5	EMPLOYEES, NONFARM-MFG
CES017	5	EMPLOYEES, NONFARM-DURABLE GOODS
CES033	5	EMPLOYEES, NONFARM-NONDURABLE GOODS
CES046	5	EMPLOYEES, NONFARM-SERVICE-PROVIDING
CES048	5	EMPLOYEES, NONFARM-TRADE, TRANSPORT, UTILITIES
CES049	5	EMPLOYEES, NONFARM-WHOLESALE TRADE
CES053	5	EMPLOYEES, NONFARM-RETAIL TRADE
CES088	5	EMPLOYEES, NONFARM-FINANCIAL ACTIVITIES
CES140	5	EMPLOYEES, NONFARM-GOVERNMENT
LHEL	2	INDEX OF HELP-WANTED ADVERTISING IN NEWSPAPERS (1967=100;SA)
LHELX	2	EMPLOYMENT: RATIO; HELP-WANTED ADS:NO. UNEMPLOYED CLF
LHEM	5	CIVILIAN LABOR FORCE: EMPLOYED, TOTAL (THOUS.,SA)
LHNAG	5	CIVILIAN LABOR FORCE: EMPLOYED, NONAGRIC.INDUSTRIES (THOUS.,SA)
LHUR	2	UNEMPLOYMENT RATE: ALL WORKERS, 16 YEARS & OVER (%,SA)
LHU680	2	UNEMPLOY.BY DURATION: AVERAGE(MEAN)DURATION IN WEEKS (SA)
LHU5	5	UNEMPLOY.BY DURATION: PERSONS UNEMPL.LESS THAN 5 WKS (THOUS.,SA)
LHU14	5	UNEMPLOY.BY DURATION: PERSONS UNEMPL.5 TO 14 WKS (THOUS.,SA)
LHU15	5	UNEMPLOY.BY DURATION: PERSONS UNEMPL.15 WKS + (THOUS.,SA)
LHU26	5	UNEMPLOY.BY DURATION: PERSONS UNEMPL.15 TO 26 WKS (THOUS.,SA)
LHU27	5	UNEMPLOY.BY DURATION: PERSONS UNEMPL.27 WKS + (THOUS,SA)
CES151	1	AVG WKLY HOURS, PROD WRKRS, NONFARM-GOODS-PRODUCING
CES155	2	AVG WKLY OVERTIME HOURS, PROD WRKRS, NONFARM-MFG
HSBR	4	HOUSING AUTHORIZED: TOTAL NEW PRIV HOUSING UNITS (THOUS.,SAAR)
HSFR	4	HOUSING STARTS:NONFARM(1947-58);TOTAL FARM & NONFARM(1959-)(THOUS.,SA
HSNE	4	HOUSING STARTS:NORTHEAST (THOUS.U.)S.A.
HSMW	4	HOUSING STARTS:MIDWEST(THOUS.U.)S.A.
HSSOU	4	HOUSING STARTS:SOUTH (THOUS.U.)S.A.
HSWST	4	HOUSING STARTS:WEST (THOUS.U.)S.A.
FYFF	2	INTEREST RATE: FEDERAL FUNDS (EFFECTIVE) (% PER ANNUM,NSA)
FYGM3	2	INTEREST RATE: U.S.TREASURY BILLS,SEC MKT,3-MO.(% PER ANN,NSA)

Table 4 continued

Name	Tran	Description
FYGM6	2	INTEREST RATE: U.S.TREASURY BILLS,SEC MKT,6-MO.(% PER ANN,NSA)
FYGT1	2	INTEREST RATE: U.S.TREASURY CONST MATURITIES,1-YR.(% PER ANN,NSA)
FYGT5	2	INTEREST RATE: U.S.TREASURY CONST MATURITIES,5-YR.(% PER ANN,NSA)
FYGT10	2	INTEREST RATE: U.S.TREASURY CONST MATURITIES,10-YR.(% PER ANN,NSA)
FYAAAC	2	BOND YIELD: MOODY'S AAA CORPORATE (% PER ANNUM)
FYBAAC	2	BOND YIELD: MOODY'S BAA CORPORATE (% PER ANNUM)
Sfygm6	1	fygm6-fygm3
Sfygt1	1	fygt1-fygm3
Sfygt10	1	fygt10-fygm3
sFYAAAC	1	FYAAAC-Fygt10
sFYBAAC	1	FYBAAC-Fygt10
FM1	6	MONEY STOCK: M1(CURR,TRAV.CKS,DEM DEP,OTHER CK'ABLE DEP)(BIL$,SA)
MZMSL	6	MZM (SA) FRB St. Louis
FM2	6	MONEY STOCK:M2(M1+O'NITE RPS,EURO$,G/P&B/DMMMFS&SAV&SM TIME DEP(BIL$,
FMFBA	6	MONETARY BASE, ADJ FOR RESERVE REQUIREMENT CHANGES(MIL$,SA)
FMRRA	6	DEPOSITORY INST RESERVES:TOTAL,ADJ FOR RESERVE REQ CHGS(MIL$,SA)
FMRNBA	6	DEPOSITORY INST RESERVES:NONBORROWED,ADJ RES REQ CHGS(MIL$,SA)
BUSLOANS	6	Commercial and Industrial Loans at All Commercial Banks (FRED) Billions $ (SA)
CCINRV	6	CONSUMER CREDIT OUTSTANDING-NONREVOLVING(G19)
PI071	6	Personal Consumption Expenditures, Price Index (2000=100), SAAR
PI072	6	Personal Consumption Expenditures-Durable Goods, Price Index (2000=100) , SA
PI073	6	Personal Consumption Expenditures-Nondurable Goods, Price Index (2000=100),
PI074	6	Personal Consumption Expenditures-Services, Price Index (2000=100), SAAR
CPIAUCSL	6	CPI All Items (SA) Fred
CPILFESL	6	CPI Less Food and Energy (SA) Fred
PCEPILFE	6	PCE Price Index Less Food and Energy (SA) Fred
PWFSA	6	PRODUCER PRICE INDEX: FINISHED GOODS (82=100,SA)
PWFCSA	6	PRODUCER PRICE INDEX:FINISHED CONSUMER GOODS (82=100,SA)
PWIMSA	6	PRODUCER PRICE INDEX:INTERMED MAT.SUPPLIES & COMPONENTS(82=100,SA)
PWCMSA	6	PRODUCER PRICE INDEX:CRUDE MATERIALS (82=100,SA)

Table 4 continued

Name	Tran	Description
PWCMSAR	5	Real PRODUCER PRICE INDEX:CRUDE MATERIALS (82=100,SA) (PWSMSAPCEPILFE)
PSCCOM	6	SPOT MARKET PRICE INDEX:BLS & CRB: ALL COMMODITIES(1967=100)
PSCCOMR	5	Real SPOT MARKET PRICE INDEX:BLS & CRB: ALL COMMODITIES(1967=100) (PSCCOMPCEPILFE)
PW561	6	PRODUCER PRICE INDEX: CRUDE PETROLEUM (82=100,NSA)
PW561R	5	PPI Crude (Relative to Core PCE) (pw561PCEPiLFE)
PMCP	1	NAPM COMMODITY PRICES INDEX (PERCENT)
EXRUS	5	UNITED STATES;EFFECTIVE EXCHANGE RATE(MERM)(INDEX NO.)
EXRSW	5	FOREIGN EXCHANGE RATE: SWITZERLAND (SWISS FRANC PER U.S.$)
EXRJAN	5	FOREIGN EXCHANGE RATE: JAPAN (YEN PER U.S.$)
EXRUK	5	FOREIGN EXCHANGE RATE: UNITED KINGDOM (CENTS PER POUND)
EXRCAN	5	FOREIGN EXCHANGE RATE: CANADA (CANADIAN $ PER U.S.$)
FSPCOM	5	S&P'S COMMON STOCK PRICE INDEX: COMPOSITE (1941-43=10)
FSPIN	5	S&P'S COMMON STOCK PRICE INDEX: INDUSTRIALS (1941-43=10)
FSDXP	2	S&P'S COMPOSITE COMMON STOCK: DIVIDEND YIELD (% PER ANNUM)
FSPXE	2	S&P'S COMPOSITE COMMON STOCK: PRICE-EARNINGS RATIO (%,NSA)
FSDJ	5	COMMON STOCK PRICES: DOW JONES INDUSTRIAL AVERAGE
HHSNTN	2	U. OF MICH. INDEX OF CONSUMER EXPECTATIONS(BCD-83)
PMI	1	PURCHASING MANAGERS' INDEX (SA)
PMNO	1	NAPM NEW ORDERS INDEX (PERCENT)
PMDEL	1	NAPM VENDOR DELIVERIES INDEX (PERCENT)
PMNV	1	NAPM INVENTORIES INDEX (PERCENT)
MOCMQ	5	NEW ORDERS (NET)-CONSUMER GOODS & MATERIALS, 1996 DOLLARS (BCI)
MSONDQ	5	NEW ORDERS, NONDEFENSE CAPITAL GOODS, IN 1996 DOLLARS (BCI)

References

Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, *88*(422), 669–679.

Armagan, A., Dunson, D. B., & Lee, J. (2013). Generalized double Pareto shrinkage. *Statistica Sinica*, *3*(1), 119–143.

Belmonte, M. A. G., Koop, G., & Korobilis, D. (2014). Hierarchical shrinkage in time-varying parameter models. *Journal of Forecasting*, *33*(1), 80–94.

Chen, Z., Iqbal, A., & Lai, H. (2011). Forecasting the probability of recessions: A probit and dynamic factor modelling approach. *Canadian Journal of Economics*, *44*(2), 651–672.

Chhikara, R., & Folks, L. (1989). *The inverse gaussian distribution: Theory, methodology, and applications*. New York: Marcel Dekker.

Christiansen, C., Eriksen, J. N., & Moller, S. T. (2014). Forecasting US recessions: The role of sentiment. *Journal of Banking and Finance*, *49*, 459–468.

De Mol, C., Giannone, D., & Reichlin, L. (2008). Forecasting using a large number of predictors: Is bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, *146*, 318–328.

Devroye, L. (1986). *Non-uniform random variate generation*. New York: Springer-Verlag.
Estrella, A., & Mishkin, F. S. (1998). Predicting U.S. recessions: Financial variables as leading indicators. *Review of Economics and Statistics*, *80*(1), 45–61.
Fernandez, C., Ley, E., & Steel, M. F. J. (2001). Benchmark priors for Bayesian model averaging. *Journal of Econometrics*, *100*, 381–427.
Fornaro P (2016) Forecasting U.S. recessions with a large set of predictors. Journal of Forecasting. doi:10.1002/for.2388.
Fossati, S. (2012). Dating U.S. business cycles with macro factors. Manuscript, University of Alberta.
Gefang, D. (2014). Bayesian doubly adaptive elastic-net Lasso for VAR shrinkage. *International Journal of Forecasting*, *30*(1), 1–11.
Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbls distribution, and the Bayesian restoration of images. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, *6*, 721–741.
George, E. I., & McCulloch, R. E. (1993). Variable Selection via Gibbs Sampling. *Journal of the American Statistical Association*, *88*, 881–889.
George, E. I., Sun, D., & Ni, S. (2008). Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics*, *142*, 553–580.
Gilks, W., Richardson, S., & Spiegelhalter, D. (1996). *Markov chain Monte Carlo in practise*. London: Chapman and Hall.
Katayama, M. (2010). Improving recession probability forecasts in the U.S. economy. Manuscript, Louisiana State University.
Koop, G., & Korobilis, D. (2012). Forecasting inflation using dynamic model averaging. *International Economic Review*, *53*, 867–886.
Korobilis, D. (2013a). Bayesian forecasting with highly correlated predictors. *Economics Letters*, *18*(1), 148–150.
Korobilis, D. (2013b). Hierarchical shrinkage priors for dynamic regressions with many predictors. *International Journal of Forecasting*, *29*(1), 43–59.
Korobilis, D. (2012). VAR forecasting using Bayesian variable selection. *Journal of Applied Econometrics*, *28*(2), 204–230.
Korobilis, D. (2016). Prior selection for panel vector autoregressions. *Computational Statistics and Data Analysis*, *101*, 110–120.
Lamnisos, D., Grin, J. E., & Steel, F. J. Mark. (2009). Transdimensional sampling algorithms for Bayesian variable selection in classification problems with many more variables than observations. *Journal of Computational and Graphical Statistics*, *18*, 592–612.
Li, Q., & Lin, N. (2010). The Bayesian Elastic Net. *Bayesian Analysis*, *5*, 151–170.
Mitchell, T. J., & Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, *83*, 1023–1036.
Panagiotelisa, A., & Smith, M. (2008). Bayesian identification, selection and estimation of semiparametric functions in high dimensional additive models. *Journal of Econometrics*, *143*, 291–316.
Park, K., & Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, *103*, 681–686.
Stankiewicz, S. (2015). Forecasting Euro area macroeconomic variables with Bayesian adaptive elastic net. Manuscript, University of Konstanz.
Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, *97*, 1167–1179.
Stock, J. H., & Watson, M. W. (2012). Generalized shrinkage methods for forecasting using many predictors. *Journal of Business and Economic Statistics*, *30*(4), 481–493.
Wright, J. H. (2006). The yield curve and predicting recessions. Finance and Economics Discussion Series, Federal Reserve Board.
Wright, J. H. (2008). Bayesian model averaging and exchange rate forecasts. *Journal of Econometrics*, *146*, 329–341.
Yuan, M., & Lin, Y. (2005). Efficient empirical Bayes variable selection and estimation in linear models. *Journal of the American Statistical Association*, *472*, 1215–1225.