Wiley StatsRef:
Statistics Reference Online

Stochastic Search Variable Selection (SSVS)

By *Konstantinos Perrakis and Ioannis Ntzoufras*

Keywords: *Bayesian variable selection, regression analysis, model search algorithms*

Abstract: The stochastic search variable selection (SSVS), introduced by George and McCulloch[1], is one of the prominent Bayesian variable selection approaches for regression problems. Some of the basic principles of modern Bayesian variable selection methods were first introduced via the SSVS algorithm such as the use of a vector of variable inclusion indicators. SSVS can effectively search large model spaces, identifying the maximum a posteriori and median probability models, and also readily produce Bayesian model averaging estimates. A number of generalizations and extensions of the method have appeared in the statistical literature implementing SSVS to a variety of applications such as generalized linear models, contingency tables, time series data, and factor analysis.

1 Introduction

One of the most prominent topics in statistical science is the selection of a model from a set of potentially plausible models under consideration. In regression analysis, this problem commonly reduces to choosing an optimal subset of variables from the set of all available covariates; see Brown[2] for a concise overview of variable selection methods. Within the Bayesian framework, variable selection is based on the evaluation of the weight of evidence quantified by Bayes factors and posterior model probabilities; see Kass and Raftery[3] and Berger[4] for more details concerning Bayes factors, Good[5] for measures of statistical evidence, and George[6] and Rice[7] for Bayesian model selection and comparison.

In regression analysis, when p predictors are available, the number of potential models is equal to 2^p, assuming that the intercept is always included and that no interactions between covariates are considered. Therefore, even for moderate values of p, the model space can be extremely large, thus rendering practically infeasible the full enumeration of all potential models and the analytic evaluation of the corresponding posterior probabilities or their approximation using information criteria such as BIC. For this reason, the development of model search algorithms, which can efficiently explore large model spaces, was of crucial importance in the early years of the Bayesian data analysis explosion, owing to the advent of Markov chain Monte Carlo (MCMC) methods in statistical science. Such algorithms should be able to detect quickly the most probable a posteriori models and deliver accurate estimates of their corresponding posterior probabilities, which can be used either for model selection or for model averaging.

A pioneering step toward this direction was achieved via the seminal paper of George and McCulloch[1], who introduced the stochastic search variable selection (SSVS) method that founded the main principles

Athens University of Economics and Business, Athens, Greece

DOI: 10.1002/9781118445112.stat07829

of modern Bayesian variable selection. SSVS was the first method appearing in the statistical literature of that time that creatively fused ideas used in hierarchical prior designs[8] and Gibbs sampling under data augmentation[9]. The main innovations of the method were: (i) the introduction of a vector of binary parameters, denoted by $\boldsymbol{\gamma}$, which was used to indicate if a variable should be included or excluded from the model (active or inactive) and (ii) that each regression coefficient was not set exactly equal to zero when a covariate was assumed to be inactive, but it was a posteriori restricted to a small neighborhood around zero via very informative zero-centered priors.

The first characteristic, that is, the binary inclusion indicators, allowed to setup a Gibbs-based algorithm for searching the model space by implementing local changes to a single covariate at a time. In this way, SSVS set the standard way of handling model uncertainty problems using MCMC algorithms. The second innovation solved an even more difficult problem concerning the implementation of MCMC methods for Bayesian model comparison. All MCMC methods until then were designed for cases where the posterior was of fixed dimension, while model selection problems involve posterior distributions of models with varying dimensions. With the simple idea of restricting inactive coefficients to small areas close to zero (instead of setting them equal to zero), George and McCulloch kept the dimension of the problem fixed across all models allowing for the standard implementation of Gibbs sampling. Thus, through SSVS one can use Gibbs sampling to explore large model spaces and directly estimate the posterior model probabilities without having to evaluate (exactly, numerically, or approximately) the marginal likelihood of each model. Moreover, Bayesian model averaging (BMA) estimates and posterior inclusion probabilities for each covariate are readily available and can be computed in a straightforward manner.

Soon after, several similar methods based on Gibbs sampling emerged, such as the Carlin and Chib[10] algorithm, the sampler of Kuo and Mallick[11], and the Gibbs variable selection method of Dellaportas *et al.*[12]. The aforementioned methods are based on similar principles and share common characteristics. An analytic review of these algorithms is outside the scope of this article as we focus on the SSVS approach; however, detailed reviews can be found in Dellaportas *et al.*[12] and more recently in O'Hara and Sillanpää[13]. Finally, Green[14] introduced the reversible-jump Metropolis–Hastings algorithm, which can be thought of as a general framework for model-search algorithms that are used in Bayesian model selection and/or averaging.

2 SSVS for the Normal Linear Model

Normal linear regression involves the setting, where we have n observations of a dependent variable Y, a set of potential predictors $X_1, X_2, \ldots, X_p$, and we assume that

$$\boldsymbol{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}) \tag{1}$$

where $\boldsymbol{Y}$ is $n \times 1$, $\mathbf{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p]$ is the $n \times p$ design matrix, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^{\mathrm{T}}$ is the vector of regression parameters, and σ^2 is a scalar. Selecting a subset of the predictors in $\mathbf{X}$ is essentially equivalent to assume that the corresponding β_j's of the predictors, which are not included in the model in Equation (1), are equal to zero; see Trader[15] for a smooth introduction in Bayesian regression.

The main characteristic of SSVS is that $\mathbf{X}$ contains all possible predictors and $\boldsymbol{\beta}$ is of fixed dimensionality p, for *all* 2^p models under consideration. Under this approach, no parameter is considered to be exactly equal to zero because a covariate is considered as "absent" or "inactive" when the corresponding parameter lies in a small "neighborhood of zero", thus, being practically negligible. This can be achieved by using a mixture of normal distributions as a prior distribution for each model coefficient β_j. Hence, the prior of each β_j given the latent binary inclusion indicator γ_j is given by the following hierarchical form

$$\beta_j|\gamma_j \sim (1-\gamma_j)N(0, \tau_j^2) + \gamma_j N(0, \tau_j^2 c_j^2) \tag{2}$$

DOI: 10.1002/9781118445112.stat07829

with

$$P(\gamma_j = 1) = 1 - P(\gamma_j = 0) = \pi_j \tag{3}$$

for $j = 1, 2, \ldots, p$. Such kind of mixture priors based on the idea of facilitating Gibbs sampling via data augmentation were initially introduced in Tanner and Wong[9]. The reasoning in the prior formulation presented in Equation (2) is to choose τ_j^2 to be "small" and $\tau_j^2 c_j^2$ to be "large" in comparison. This way, when $\gamma_j = 1$, β_j is present in the model with a prior distribution that is vague so that the posterior distribution will be mainly determined by the data. In contrast, when $\gamma_j = 0$, β_j is considered to be absent from the model, and the prior becomes more concentrated around the null hypothesis, forcing this parameter to be shrunk toward zero. In a general multivariate form, the prior in Equation (2) can be expressed as

$$\boldsymbol{\beta}|\boldsymbol{\gamma} \sim N_p(\mathbf{0}\,, \mathbf{D}_{\boldsymbol{\gamma}}\mathbf{R}\mathbf{D}_{\boldsymbol{\gamma}}) \tag{4}$$

where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_p)^{\mathrm{T}}$, $\mathbf{D}_\gamma \equiv \mathrm{diag}(\alpha_1\tau_1, \alpha_2\tau_2, \ldots, \alpha_p\tau_p,)$ with $\alpha_j = 1$ if $\gamma_j = 0$ and $\alpha_j = c_j$ if $\gamma_j = 1$, and $\mathbf{R}$ is the prior correlation matrix that can be defined accordingly (see next section).

The conditional prior of β_j given that it is "inactive" (i.e., $\gamma_j = 0$) is the one that essentially restricts the posterior distribution of β_j to lie in a "small" area around zero instead of setting it exactly equal to zero. This is implemented using a "ridge regression" type of shrinkage. This property is also the main feature that differentiates SSVS from the other Gibbs-based model search algorithms such as the ones proposed by Kuo and Mallick[11] and Dellaportas *et al.*[12]. Moreover, the SSVS posterior probabilities and Bayes factors will not be the same as the ones that result from using equality-to-zero constraints as the underlying models are slightly different. Nevertheless, the two approaches will tend to converge to the same results as the prior variance of the inactive effects approaches zero. The "spike and slab" prior of Mitchell and Beauchamp[8] can be considered as an ancestor of SSVS as the two approaches share some common ideas.

In the above-mentioned model formulation, we have not discussed about the constant term that is usually included in all regression models. Without loss of generality, we do not need to treat the constant term separately but as a simple covariate with all values equal to one, that is, $\boldsymbol{X}_1 = (1, 1, \ldots, 1)^T$. In such case, we can retain the constant term in all models by simply specifying the corresponding prior probability of inclusion equal to one, that is, $\pi_1 = 1$. Equivalently, the constant can be eliminated from the model formulation without changing the interpretation of the model coefficients if we center both the response and the explanatory variables at their sample means.

SSVS is completed by specifying the prior distributions of σ^2 and $\boldsymbol{\gamma}$. For the variance component, the most convenient option is the conjugate inverse-gamma prior, that is

$$\sigma^2|\boldsymbol{\gamma} \sim \mathrm{IG}(\nu_{\boldsymbol{\gamma}}/2, \nu_{\boldsymbol{\gamma}}\lambda_{\boldsymbol{\gamma}}/2) \tag{5}$$

As noted in George and McCulloch[1], conditioning upon $\boldsymbol{\gamma}$ provides flexibility, in the sense that one can incorporate dependency between $\boldsymbol{\beta}$ and σ^2; for instance, one may want to allow the variance to decrease as the dimension of $\boldsymbol{\beta}$ increases. Finally, for $\boldsymbol{\gamma}$, the authors suggest that a reasonable option is to consider the γ_j's to be independent so that

$$\pi(\boldsymbol{\gamma}) = \prod_{j=1}^{p} \pi_j^{\gamma_j}(1 - \pi_j)^{(1-\gamma_j)} \tag{6}$$

The prior in Equation (6) implies that the inclusion of X_ℓ is independent of the inclusion of X_j for all $\ell \neq j$. This prior is a standard option that facilitates Gibbs sampling. Nevertheless, alternative prior forms that allow incorporating structural information about the design matrix are also available; see Section 5 for some details.

Given the priors in Equations (4)–(6), the corresponding full conditional distributions of $\boldsymbol{\beta}$, σ^2, and $\boldsymbol{\gamma}$ are all known in closed form (normal for $\boldsymbol{\beta}$, inverse-gamma for σ^2, and Bernoulli for γ_j), allowing for fast and efficient Gibbs sampling; see George and McCulloch[1] for details. Because of its simplicity, SSVS can also be used in cases where the marginal likelihood can be evaluated analytically but the model space is large, that is, as a better alternative to full enumeration (which may not be feasible) or to an MC^3 algorithm (see

DOI: 10.1002/9781118445112.stat07829

Ref. 3, for details), both of which involve inversion of (possibly) large matrices depending on the dimension of the models under consideration. Finally, SSVS can be easily implemented in WinBUGS and other related software (e.g., OpenBUGS or JAGS); see Refs 16 and 17 for detailed examples.

It is worth noting that in subsequent work, George and McCulloch[18] also proposed an alternative prior for $\boldsymbol{\beta}$, which is further conditioned upon σ^2, namely $\boldsymbol{\beta}|\sigma^2, \boldsymbol{\gamma} \sim N_p(\mathbf{0}, \sigma^2\mathbf{D}_{\boldsymbol{\gamma}}\mathbf{R}\mathbf{D}_{\boldsymbol{\gamma}})$. This prior on $\boldsymbol{\beta}$, combined with the priors in Equations (5) and (6), results in a conjugate design that simplifies calculations as the only requirement in this case is to evaluate $\pi(\boldsymbol{\gamma}|\boldsymbol{Y})$, which is known up to a proportionality constant; this can done analytically for small or moderate p and through MCMC methods for large p; see George and McCulloch[18] for details.

The output of SSVS is a posterior sample $\{\boldsymbol{\beta}^{(t)}, \sigma^{(t)}, \boldsymbol{\gamma}^{(t)}\}$ for $t = 1, \ldots, T$. Focus is placed on $\boldsymbol{\gamma}$ as the relative frequencies of all sampled combinations of $\boldsymbol{\gamma}$ will provide estimates of the posterior model probabilities. The posterior inclusion probability of each covariate X_j can be easily estimated as $\widehat{P}(\gamma_j = 1|\boldsymbol{Y}) = T^{-1}\sum_{t=1}^{T}\gamma_j^{(t)}$. Furthermore, for any quantity of interest θ, the BMA estimate is simply the sample mean $\overline{\theta}$ obtained from the SSVS output as $\theta^{(t)}$ is generated at each iteration t from the full conditional posterior distribution of model $\boldsymbol{\gamma}^{(t)}$.

3 Specification of Hyperparameters

From the simple version of SSVS, which uses independent priors in Equation (2), we can obtain results comparable to the ones obtained from standard Bayesian variable selection methods that assume $\beta_j = 0$ for nonimportant effects and a normal $N(0, \sigma^2_{\beta_j})$ prior when $\beta_j \neq 0$. That occurs when the “large” prior variance $c_j^2\tau_j^2$ of SSVS is set equal to $\sigma^2_{\beta_j}$ and τ_j^2 is chosen to be suitably low. Note that for $\tau_j^2 \to 0$, the underlying posterior model probabilities of the two approaches will coincide, however, in this case SSVS will be less and less efficient and mobile in model space. Therefore, τ_j^2 must be tuned carefully in order to keep SSVS efficient and at the same time have a good proxy for posterior model probabilities of variable selection based on point-zero null hypotheses.

For the specification of τ_j^2 and c_j^2, George and McCulloch[1] describe a “semiautomatic” procedure in order to select reasonable values. The idea relies first on the fact that c_j can be interpreted as the odds ratio of excluding X_j when β_j is very close to zero and second on fine-tuning the ratio of statistical over practical significance measured by the quantity $\widehat{\sigma}_{\beta_j}/\tau_j$, with $\widehat{\sigma}_{\beta_j}$ being the observed standard error of the least squares estimate of β_j. On the basis of sensitivity analyses in George and McCulloch[1], fixed values of $\widehat{\sigma}_{\beta_j}/\tau_j$ and c_j set to (1,5), (10,100) usually perform well under most situations, although some caution may be needed depending on the characteristics of the data set, such as sample size, number of predictors, and so forth. Guidelines for specifying these parameters using logical arguments based on the size of the coefficient can be found in Ntzoufras[16] for log-linear models for contingency tables and in Mavridis and Ntzoufras[19] for the loadings in factor analysis models.

Regarding the prior correlation matrix $\mathbf{R}$, two prior choices, representing two extremes, are of particular interest; namely, $\mathbf{R} = \mathbf{I}$, which means that the components in $\boldsymbol{\beta}$ are considered independent a priori, and $\mathbf{R} \propto (\mathbf{X}^T\mathbf{X})^{-1}$, which corresponds to the g-prior of Zellner[20], where the prior correlation is identical to the design correlation multiplied by a scalar.

For parameter σ^2, prior ignorance can be expressed through the choice $\nu_{\boldsymbol{\gamma}} \equiv 0$, and any value of $\lambda_{\boldsymbol{\gamma}}$ as in this case the prior in Equation (5) does not contribute any information to the posterior. Alternatively, one can regard $[\nu_{\boldsymbol{\gamma}}/(\nu_{\boldsymbol{\gamma}} - 2)]\lambda_{\boldsymbol{\gamma}}$ as a prior estimate of σ^2 (from an imaginary experiment of $\nu_{\boldsymbol{\gamma}}$ data points) and define this quantity as a decreasing function of $p_{\boldsymbol{\gamma}} = \sum_{j=1}^{p}\gamma_j$, so that higher-dimensional models will be a priori expected to have a lower residual variance.

Lastly, for the product-Bernoulli distribution in Equation (6), assigned to $\boldsymbol{\gamma}$, a usual choice is the uniform prior, where each $\boldsymbol{X}_j$ has an equal chance ($\pi_j = 0.5$) of being included or excluded from the model, so that $\pi(\boldsymbol{\gamma}) \equiv 2^{-p}$. Alternatively, one can use mixtures of beta-binomial priors that result in marginal priors of the form $\pi(\boldsymbol{\gamma}) = w_{p_\gamma} \binom{p}{p_{\boldsymbol{\gamma}}}^{-1}$, where w_{p_γ} is a prior weight for a model of size $p_{\boldsymbol{\gamma}}$. A typical option is $w_{p_\gamma} = 1/(p+1)$, which results from a beta mixture distribution with parameters a and b equal to one, that is, a uniform prior on π_j; see Scott and Berger[21].

4 Extensions of SSVS

Ever since the initial development of SSVS for normal linear models, the applicability of the method has been extended to a wide variety of models. George *et al.*[22] considered SSVS for generalized linear models (GLMs) and demonstrated its use for probit regression. Brown *et al.*[23] extended SSVS to multivariate normal regression providing fast and efficient algorithms for problems that may involve a large number of predictors. Ntzoufras *et al.*[24] focused on log-linear models for multiway contingency tables. This special case of GLMs has some features, which make implementation of SSVS challenging; specifically, (i) for problems involving factors with more than two levels the terms for main and interaction effects are represented by more than one parameter, so that the prior in Equation (2) can be multivariate and (ii) the γ_j's should not be considered independent a priori, as in Equation (6), owing to the fact that values of $\boldsymbol{\gamma}$ corresponding to nonhierarchical models are prohibited. Further extensions of SSVS have been developed for factor analysis models, where the goal is to interpret and quantify the interrelationships between observed variables (items) and latent variables (factors). Within the factor analytic framework, Dunson[25] introduced the stochastic search factor selection (SSFS) approach, while Mavridis and Ntzoufras[19] further generalized the methodology by developing models and algorithms for stochastic search of item selection (SSIS) as well as stochastic search factor and item selection (SSFIS). Further extensions and modifications of SSVS are used in genetics, especially in gene-mapping applications Refs 26–28, in time series and econometric models Refs 29 and 30 and, recently, in quantile regression[31].

5 SSVS: Toward the Future

Ročková and George[32] have recently introduced the expectation-maximization variable selection (EMVS) algorithm. EMVS is actually founded on the hierarchical model specification of SSVS assuming ridge type of shrinkage for the "inactive" covariates and independent mixtures of normals for the model coefficients. These are the main features that make the EM algorithm applicable in the Bayesian variable formulation. If the sharp point-null restriction to zero is assumed or multivariate-dependent priors are adopted, then the EM cannot be implemented (at least in an obvious and efficient way) as computations require calculations over the whole model space. EMVS is naturally much faster than its stochastic counterpart and is especially suited for high-dimensional problems with large p or even $p > n$. Especially for the latter situation, EMVS is particularly flexible as it allows incorporating shrinkage hyper-priors for the "slab" part of the hierarchical prior in Equation (2). Furthermore, structured information about the design matrix can be naturally incorporated into EMVS through the use of independent logistic regression priors or Markov random field priors for $\boldsymbol{\gamma}$ instead of the product Bernoulli prior in Equation (6). According to Ročková and George[32], future directions include, among others, extending the EMVS methodology to GLM's, Gaussian graphical models, factor analysis models, and multivariate regression models.

Related Articles

Bayes Factors; **Bayesian Model Selection**; **Regression, Bayesian**; **Statistical Evidence**; **Bayesian Methods for Model Comparison**; **Variable Selection**.

References

[1] George, E.I. and McCulloch, R.E. (1993) Variable selection via Gibbs sampling. *J. Am. Stat. Assoc.*, **85**, 398–409.

[2] Brown, P.J. (2014) *Variable Selection*, Wiley StatsRef: Statistics Reference Online.

[3] Kass, R.E. and Raftery, A.E. (1995) Bayes factors and model uncertainty. *J. Am. Stat. Assoc.*, **90**, 773–795.

[4] Berger, J.O. (2014) *Bayes Factors*, Wiley StatsRef: Statistics Reference Online.

[5] Good, I.J. (2014) *Statistical Evidence*, Wiley StatsRef: Statistics Reference Online.

[6] George, E.I. (2014) *Bayesian Model Selection*, Wiley StatsRef: Statistics Reference Online.

[7] Rice, K. (2014) *Bayesian Methods for Model Comparison*, Wiley StatsRef: Statistics Reference Online.

[8] Mitchell, T.J. and Beauchamp, J.J. (1988) Bayesian variable selection in linear regression (with discussion). *J. Am. Stat. Assoc.*, **83**, 1023–1036.

[9] Tanner, M.A. and Wong, W. (1987) The calculation of posterior distributions by data augmentation (with discussion). *J. Am. Stat. Assoc.*, **82**, 528–550.

[10] Carlin, B.P. and Chib, S. (1995) Bayesian model choice via Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B*, **57**, 473–484.

[11] Kuo, L. and Mallick, B. (1998) Variable selection for regression models. *Sankya B*, **60**, 65–81.

[12] Dellaportas, P., Forster, J.J., and Ntzoufras, I. (2002) On Bayesian model and variable selection using MCMC. *Stat. Comput.*, **12**, 27–36.

[13] O'Hara, R.B. and Sillanpää, M.J. (2009) A review of Bayesian variable selection methods: what, how and which. *Bayesian Anal.*, **4**, 85–118.

[14] Green, P.J. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 711–732.

[15] Trader, R.L. (2014) *Regression, Bayesian*, Wiley StatsRef: Statistics Reference Online.

[16] Ntzoufras, I. (2002) Gibbs variable selection using BUGS. *J. Stat. Softw.*, **7** (7), 1–19.

[17] Ntzoufras, I. (2009) *Bayesian Modeling Using WinBUGS*, Wiley Series in Computational Statistics, John Wiley & Sons, Inc., Hoboken, NJ.

[18] George, E.I. and McCulloch, R.E. (1997) Approaches for Bayesian variable selection. *Stat. Sin.*, **7**, 339–373.

[19] Mavridis, D. and Ntzoufras, I. (2014) Stochastic search item selection for factor analytic models. *Br. J. Math. Stat. Psychol.*, **67**, 284–303.

[20] Zellner, A. (1986) On assessing prior distributions and Bayesian regression analysis with g-prior distributions, in *Bayesian Inference and Decision Techniques: Essays in Honour of Bruno de Finetti* (eds P. Goel and A. Zellner), North-Holland, Amsterdam, pp. 233–243.

[21] Scott, J.G. and Berger, J.O. (2010) Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. *Ann. Stat.*, **38**, 2587–2619.

[22] George, E.I., McCulloch, R.E., and Tsay, R.S. (1996) Two approaches to Bayesian model selection with applications, in *Bayesian Analysis in Statistics and Econometrics: Essays in Honour of Arnold Zellner* (eds D.A. Berry, K.M. Chaloner, and J.K. Geweke), John Wiley & Sons, Inc., New York, pp. 339–348.

[23] Brown, P.J., Vannucci, M., and Fearn, T. (1998) Multivariate Bayesian variable selection and prediction. *J. R. Stat. Soc. Ser. B*, **60**, 627–641.

[24] Ntzoufras, I., Forster, J.J., and Dellaportas, P. (2000) Stochastic search variable selection for log-linear models. *J. Stat. Comput. Simul.*, **68**, 23–37.

[25] Dunson, D.B. (2006) Efficient Bayesian model averaging in factor analysis. ISDS Discussion Paper, Institute of Statistics and Decision Sciences, Duke University.

[26] Yi, N., George, V., and Allison, D. (2003) Stochastic search variable selection for identifying multiple quantitative trait loci. *Genetics*, **164**, 1129–1138.

[27] Meuwissen, T.H.E. and Goddard, M.E. (2004) Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data. *Genet. Sel. Evol.*, **36**, 261–279.

[28] Swartz, M.D., Kimmel, M., Mueller, P., and Amos, C.I. (2006) Stochastic search gene suggestion: a Bayesian hierarchical model for gene mapping. *Biometrics*, **62**, 495–503.

[29] George, E.I., Sun, D., and Ni, S. (2008) Bayesian stochastic search for VAR model restrictions. *J. Econometrics*, **142**, 553–580.

[30] Jochmann, M., Koop, G., and Strachan, R.W. (2010) Bayesian forecasting using stochastic search variable selection in a VAR subject to breaks. *Int. J. Forecast.*, **26**, 326–347.

[31] Yu, K., Chen, C.W.S., Reed, C., and Dunson, D.B. (2013) Bayesian variable selection in quantile regression. *Stat. Intreface*, **6**, 261–274.

[32] Ročková, V. and George, E.I. (2014) EMVS: the EM approach to Bayesian variable selection. *J. Am. Stat. Assoc.*, **109**, 828–846.

DOI: 10.1002/9781118445112.stat07829