Variational Bayesian Approach for Vector Autoregression Model Learning

Wei-Ting Lai and Ray-Bing Chen

November 18, 2020

Variational Bayesian Approach for Vector Autoregression Model Learning

Wei-Ting Lai
Department of Statistics,
National Cheng Kung University
Taiwan
R28051020@gs.ncku.edu.tw

Ray-Bing Chen
Department of Statistics & Institute of Data Science ,
National Cheng Kung University
Taiwan
rbchen@ncku.eud.tw

***Abstract*—The vector autoregressive (VAR) model is one of the cores of analyzing the structure of multivariate time series over time. VAR is becoming more and more popular with complex data structure and huge data size. However, at the same time, the traditional MCMC application on the VAR model encounters the problem of excessive calculation time. To overcome the problem, we proposed the variational Bayesian Method for the VAR data analysis. The performance of the proposed method is illustrated via simulation studies.**

Keywords—VAR model, variable selection, Variational Bayes

I. Introduction

The vector autoregressive (VAR) model is one of the cores for analyzing the time-dependent structure of multivariate time series. It is an extension of the univariate autoregressive (AR) model, which not only has sequence dependence in each time series but also has mutual dependence in different time series. In recent years, due to the complicate data structure and huge data size, VAR is more and more popular. For example, Weron [1] and Ziel and Weron [2] use VAR model in the electricity price related applications. However, with the increasing of the dimensionality, the VAR model has the overparametric problem which would give some troubles in the data analysis, because when the model has m time series and each time series has p lags, there should estimate pm^2 coefficients in the VAR model.

To overcome this weakness in VAR model, we may impose some meaningful structural assumptions and then the structure selection is implemented to reduce the dimensionality of the parameters. Bickel and Song [3] introduced three different parameter structures and a LASSO type method was proposed to identify the proper structures. Nicholson et al. [4] generalized their works to cover more VAR structures and in addition to LASSO penalty function, they also considered the group and sparse group penalties. Instead of penalty approach for structure selection, Bayesian approach is also commonly used. Chu et al. [5] proposed a Bayesian structure selection approach to deal with three VAR structures mentioned in Bickel and Song [3]. In Chu et al. [5], the indicators are added into the model to denote the meaningful VAR structures and then an MCMC algorithm is used to generate the posterior samples of the indicators for the future inference. They did show the advantage of their Bayesian approach via several simulations and a real example.

Due to the nature of MCMC algorithms, as model complexity and the sample size increase, the computing time increases dramatically. For example, in Chu et al. [5], their approach would take 1 to 6 hours for the case of 20 dimensional simulation data according the group structures. Thus the goal of this work is to increase the computational efficiency for the VAR analysis approach.

II. Method

To increase the computational efficiency of the Bayesian inference approach, the variational Bayesian approach, proposed by Titsias and Lazaro Gredilla [6] and Carbonetto and Stephens [7], is adopted here. Instead of directly generating posterior samples, the variational Bayesian method is to find the best approximate distribution of the true posterior by minimizing Kullback-Leibler divergence (KL-divergence) for Bayesian inference. Thus according to [8], in the variational Bayesian approach, the key is to solve the following optimization problem,

$$\min_{Q\in\mathbb{Q}}\left(-E^{Q}log\frac{P}{Q}\right)=-\int Qlog\frac{P}{Q}dQ,$$

where P is used to denote the true posterior distribution and Q is the approximate distribution. Consider the structure selection problems. Cai et al. [9] proposed the sparse group variable selection of the variational Bayesian method with spike-and-slab prior in the linear model. In the prior, using spike-and-slab prior to deal with the variable selection problems. Thus, the goal of this work is to propose the variational Bayesian approach for the VAR modeling according to the idea of Cai et al. [9].

In this work, following Song and Bickel [3] and Chu et al. [5], the three structures for the parameter matrix in VAR model are considered. There are universal grouping, sigmentize grouping and no grouping. The universal group is that all columns are as one group in the same row. For the segmentized grouping structure, all of the time series is divided into nonoverlapping group sets by prior knowledge. Then, coefficients are estimated by segment-by-segment. Finally, for the no grouping structure, each time series is viewed as individual and estimates it column-by-column. In our proposed variational Bayesian approach, similar to Chu et al. [5], latent variables are augmented into the VAR model to denote the active structures. Then based on the binomial priors for the latent variables, the independent spike-and-slap priors are set for the parameters. Due to these prior assumptions, the approximation posterior function, Q, is defined as the product of these prior densities. Finally, the corresponding minimization problem is solved via an expectation-maximization (EM) type method.

III. Simulation

To illustrate the performance of the proposed variational Bayesian approach, VARVB, simulation studies in Chu et al.

[5] are considered. In this section, three scenarios with respect to the three different structures are set up and the dimension of time series is 10 and 20 respectively. In each scenario, the variance structures of the error terms come from 2 different set-ups, the identity matrix, I, and the covariance matrix Σ respectively. We generate data with 301 samples and use 300 samples for model training with a fixed number of lags, 10, and the last single sample for prediction purpose. Overall, 10 replicates are implemented for each scenario by independently re-generating the data and the number of groups is assumed to be known in this simulation. In addition to VARVB, we also implement the VAGSA method in Chu et al. [5] for the comparison purpose. In the VAGSA, it iterated 3,000 sweeps and make inferences from the last 1,000 sweeps. For performance comparisons, we report four measurements, the true positive rate (TPR); the false positive rate (FPR); the average of the model sizes ($\widehat{M}$); and the average (standard deviation) of the mean square prediction error ($\overline{MSPE}$). As a result, in most scenarios, the TPR and $\overline{MSPE}$ in both methods and cases are similar. In the FPR and $\widehat{M}$, our method is better than VAGSA. However, in the cases of segmentize grouping (scenario 2) with 20-dimension time series with covariance error, do not work well for our method. The reason may be due to that in one of 10 replicates, our method selects fewer correct variables and more incorrect variables which causes a bad result. In the elapsed time, our method has a good reduction average 4 times the time cost in all of the cases. especially in the case with group structure. It can reduce 10 times than VASGA. Thus, it means that we can reduce the time and without losing a lot of accuracies.

IV. CONCULSION

This research proposes a priori variational Bayesian method for a VAR model. Due to our simulation results, overall the performance of the proposed method is comparable with the traditional Bayesian approach. But this new variational Bayesian approach does save a lot of computation cost and have potential to deal with the larger data sizes. We leave this as a future work.

TABLE I. THE RESULT IN 10 DIMENSION VAR MODEL

	Method	TPR	FPR	$\widehat{M}$	$\overline{MSPE}$
Scenario 1				True=72	
I_{10}	**VARVB**	100%	0.06%	72.60	0.92(0.23)
	VAGSA	100%	0.34%	74.20	0.94(0.25)
Σ_{10}	**VARVB**	100%	0.06%	72.60	0.92(0.84)
	VAGSA	100%	0.23%	74.20	0.87(0.86)
Scenario 2				True=40	
I_{10}	**VARVB**	100%	0.13%	41.20	0.90(0.21)
	VAGSA	100%	0.49%	44.70	0.94(0.26)
Σ_{10}	**VARVB**	100%	0.03%	40.30	0.73(0.36)
	VAGSA	99%	0.46%	44.40	0.88(0.39)
Scenario 3				True=18	
I_{10}	**VARVB**	97%	0.18%	19.30	0.88(0.21)
	VAGSA	96%	0.11%	18.40	0.87(0.20)
Σ_{10}	**VARVB**	96%	0.13%	18.69	0.72(0.37)
	VAGSA	96%	0.09%	18.20	0.87(0.37)

TABLE II. THE RESULT IN 20 DIMENSION VAR MODEL

	Method	TPR	FPR	$\widehat{M}$	$\overline{MSPE}$
Scenario 1				True=145	
I_{20}	**VARVB**	100%	0.03%	146.00	1.07(0.18)
	VAGSA	100%	0.18%	152.10	1.08(0.16)
Σ_{20}	**VARVB**	100%	0.01%	145.10	0.92(0.44)
	VAGSA	100%	0.14%	150.00	0.88(0.44)
Scenario 2				True=109	
I_{20}	**VARVB**	100%	0.06%	111.20	1.09(0.14)
	VAGSA	100%	0.22%	117.60	1.07(0.13)
Σ_{20}	**VARVB**	99%	0.54%	129.30	0.75(0.20)
	VAGSA	100%	0.17%	116.00	0.88(0.20)
Scenario 3				True=27	
I_{20}	**VARVB**	99%	0.07%	29.30	1.06(0.14)
	VAGSA	99%	0.13%	32.10	1.06(0.14)
Σ_{20}	**VARVB**	98%	0.06%	28.70	0.84(0.34)
	VAGSA	99%	0.10%	30.60	0.89(0.34)

TABLE III. ELAPSED TIME OF SIMULATION IN SECONDS

m=10	**VARVB**	**Scenario 1**	**Scenario 2**	**Scenario 3**
CPU times (I_{10})		83.17	200.89	786.38
CPU times (Σ_{10})		104.75	296.90	1154.94
	VAGSA	**Scenario 1**	**Scenario 2**	**Scenario 3**
CPU times (I_{10})		1225.66	2835.45	5294.96
CPU times (Σ_{10})		1096.47	1685.83	4556.46
m=20	**VARVB**	**Scenario 1**	**Scenario 2**	**Scenario 3**
CPU times (I_{20})		221.46	671.62	3227.11
CPU times (Σ_{20})		359.47	1475.56	4434.28
	VAGSA	**Scenario 1**	**Scenario 2**	**Scenario 3**
CPU times (I_{20})		2483.74	5024.19	14935.81
CPU times (Σ_{20})		2586.12	4790.69	21157.75

REFERENCES

[1] R. Weron, "Electricity price forecasting: A review of the state-of-the-art with a look into the future," International journal of forecasting, vol. 30, pp. 1030-1081, 2014.

[2] F. Ziel, and R.Weron, " Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks, " Energy Economics, vol. 70, pp. 396-420, 2018.

[3] S. Song, and P. J. Bickel, " Large vector auto regressions, " preprint 2011 .

[4] W. B. Nicholson, D. S. Matteson, and J. Bien, " Structured regularization for large vector autoregressions, " Cornell University, 2014.

[5] C. H. Chu, M. N. Lo Huang, S. F.Huang and R. B. Chen, " Bayesian structure selection for vector autoregression model, " Journal of Forecasting, vol. 38, pp. 422-439, 2019.

[6] M. K. Titsias, and M. Lázaro-Gredilla, " Spike and slab variational inference for multi-task and multiple kernel learning, " In Advances in neural information processing systems, pp. 2339-2347, 2011.

[7] P. Carbonetto, and M. Stephens, " Scalable variational inference for Bayesian variable selection in regression, and its accuracy in genetic association studies, " Bayesian analysis, vol. 7, pp. 73-108. 2012.

[8] C. M. Bishop. Pattern recognition and machine learning. Springer, 2006.

[9] M. Cai, M. Dai, J. Ming, H. Peng, J. Liu, and C. Yang, " BIVAS: a scalable Bayesian method for bi-level variable selection with applications, " Journal of Computational and Graphical Statistics, vol 29, pp. 40-52 , 2020