

Full Length Research Paper

Model selection on tourism forecasting: A comparison between Bayesian model averaging and Lasso

Jiyuan Wang¹, Geng Peng^{1*} and Shouyang Wang²

¹School of Economics and Management, University of Chinese Academy of Sciences, China.

²Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China.

Received 29 December, 2016; Accepted 3 February, 2017

This study tries to tackle the tourism forecasting problem using online search queries. This recent-developed methodology is subject to several criticisms, one of which is how to choose satisfying search queries to be built in the forecasting model. This study compares two popular candidates, which are the Bayesian Model Averaging (BMA) approach and the Least Absolute Shrinkage and Selector Operator (Lasso) approach. Evidence shows that the two approaches produce similar forecasting performance but different query selection results.

Key words: Tourism forecasting, query selection, Bayesian model averaging, Lasso, Baidu query data.

INTRODUCTION

In recent years, much attention on forecasting based on user generated content (UGC) has been paid. One representative is the data from search engine services. The users of search engines search certain key words, or queries, to acquire some information on purpose, which leaves the service providers a new route to analysis users' motivation and information demand.

In addition to these analyses, forecasting on certain issues can be conducted. A typical example is the Google Flu Trends (GFT), initialized in 2008, which has made several successful forecasts of flu outbreaks in US. The first paper on this issue is Ginsberg et al. (2009).

However, GFT is not as successful as it was initialized, and it has gone through many criticisms. Those criticisms come from the worries about the uncertainties of the query data generation process (DGP) and the model

instability (Butler, 2013) for example. Nevertheless, the forecasting on other aspects based on query data still contributes knowledge to both practitioners and academia, for example in tourism forecasting (Bangwayo-Skeete and Skeete, 2015; Yang et al., 2015; Li et al., 2017). The difference of the performance in different areas using search query data emerges partially from the variable selection approaches and forecasting models.

In most research, variable selection procedures based on correlation coefficient, purely subjective index compositing or other criteria might be problematic. Several researches try to improve the forecasting model by introducing variable selection models, such as spike and slab (Scott and Varian, 2014) and other pioneer approaches. As pointed by Brynjolfsson et al. (2015),

*Corresponding author. E-mail: penggeng@ucas.ac.cn.

successful forecasting based on query data has several problems to be settled, one of which is to solve the variable selection problem.

This study employed two frequently-used models with variable selection property to build the forecasting model on tourists forecasting. They are the Bayesian Model Averaging (BMA) approach and the Least Absolute Shrinkage and Selector Operator (Lasso) approach. Lasso is first proposed by Tibshirani (2011), and has many successful variants in predicting genome, economic amounts etc. BMA is sufficiently described in Hoeting et al. (1999), and it also has several follow-up updates.

In general, the two approaches can both select variables and conduct forecasting; however the mechanisms are totally different. The BMA approach, presumably, believes uncertainties exist in each model. Instead of choosing the best model, it advocates to combine those top models selected through Bayesian Factor or other criteria. Lasso applies soft-thresholding by introducing a penalty on model parameters when minimizing its squared error objective. Due to the different originalities of the two approaches, the theoretical comparison becomes extremely hard and, even, trivial.

However, from a practitioner's point of view, choosing between the two approaches is extremely useful. Except for the forecasting performance, the best model sizes and which variables are selected through BMA and Lasso also contribute to the forecasting practitioners.

Forecasting models

This section is to recap the main specification of the Bayesian Model Averaging (BMA), Least Absolute Shrinkage and Selector Operator (Lasso). Without loss of generality, we assume the underlying true model is of linear form, as shown in Equation 1:

$$y = \alpha + X\beta + \epsilon, \epsilon \sim N(0, \sigma^2 I), \quad (1)$$

where α is the constant, X contains K covariates with N observations, and β is the $K \times 1$ vector. Besides, in this model, we assume homoscedastic standard errors. Based on this specification, we now discuss the main results of BMA and Lasso.

Bayesian model averaging

To allow model uncertainties, we rearrange model (1) as

$$y = \alpha_\gamma + X_\gamma \beta_\gamma + \epsilon, \epsilon \sim N(0, \sigma^2 I), \quad (2)$$

with γ as the model index, where $\gamma = 1, 2, \dots, 2^K$. Each model of (2) in the Bayesian framework requires assumptions on priors of coefficient parameters β and variance σ^2 . Following the tradition assumption,

(Feldkircher, 2012; Moral-Benito, 2015), we assume β follows a normal distribution with parameter Zellner's g in Feldkircher (2012):

$$\beta_\gamma | g \sim N\left(0, \sigma^2 \left(\frac{1}{g} X'X\right)^{-1}\right), \quad (3)$$

where different specifications of Zellner's g would yield different model performance and model size. The choice of g could be fixed, for example, N (Unit Information Prior), empirical Bayes g whose idea is to use the information in the data to help improve the models, and Hyper- g prior who gives the ratio $\frac{g}{1+g}$ a Beta prior and usually we set $g = N$ therefore $E \frac{g}{1+g} = \frac{N}{1+N}$. Importantly, $\frac{g}{1+g}$ acts as a shrinkage factor because from Equation (3) we have

$$E(\beta_\gamma | y, X, g, M_\gamma) = \frac{g}{1+g} \widehat{\beta}_\gamma^{OLS}, \quad (4)$$

hence smaller g yields harder shrinkage on parameter estimates, which can be regarded as a way of variable selection.

The variance σ^2 has a non-informative prior

$$p(\sigma) \propto \sigma^{-1}. \quad (5)$$

In addition, each model M_γ is assigned with a prior probability $p(M_\gamma)$. Normally, the priors of the models can be of any form, as long as it follows the subjective knowledge of the person who conducts forecasting. The most convenient priors are the Uniform, the Binomial and the Beta-Binomial (Doppelhofer and Miller, 2004; Ley and Steel, 2009). After specifying all the priors, we can then calculate the posterior model inclusion probabilities through

$$p(M_\gamma | y, X) = \frac{p(y | M_\gamma, X) p(M_\gamma)}{p(y | X)}, \quad (6)$$

where $p(y | X)$ is the integrated likelihood, and it can be calculated through

$$p(y | X) = \sum_{s=1}^{2^K} p(y | M_s, X) p(M_s) \quad (7)$$

After estimating the in-sample posterior densities of models and parameters, we can then conduct out-of-sample forecasting y_f according to

$$p(y_f | y) = \sum_{i=1}^{2^K} p(M_i | y) \int p(y_f | \theta_i, y, M_i) p(\theta_i | y, M_i) d\theta_i, \quad (8)$$

where $\theta_i = ((\alpha_i, \beta_i)')$, while $p(y_t|\theta_i, y, M_i)$ is the out-of-sample likelihood and $p(\theta_i|y, M_i)$ is the posterior of parameters in model i . When the number of regressors becomes larger, the enumerate methods seem to be implausible, instead MCMC sample is the best choice to approximate the posterior densities of models. At any state, for instance t , staying at model i will jump to model j at state $t + 1$ according to the transition density p_{ij} . However, in general, the transition density is neither known nor reversible. Therefore, the most frequently-used algorithm is the Metropolis-Hastings algorithm in this case and we adopt a random walk kernel during the sampling, the transition from model i to model j is based on

$$p_{ij} = \min \left\{ 1, \frac{p(M_j|y, X)}{p(M_i|y, X)} \right\}, \quad (9)$$

then we can approximate all the relevant posteriors we need.

Lasso

The Lasso estimator on model (1) is to minimize the following criterion function:

$$(y - X\beta)'(y - X\beta) + \lambda \sum_{j=1}^K |\beta^j|, \quad (10)$$

and Lasso has a nice property against Ridge regression estimator

$$(y - X\beta)'(y - X\beta) + \lambda \sum_{j=1}^K (\beta^j)^2, \quad (11)$$

is that Lasso exhibit hard thresholding property which can rule out some irrelevant variables stricter than Ridge estimator.

Commonly, to guarantee the model stability of Lasso such that we can also mitigate model uncertainty to some degree, we apply cross-validation. Basically, we divide the sample into d subsamples. By taking $d - 1$ out of d subsample we estimate Lasso and take the rest one to compute the prediction errors. Then taking the average of the results, we can have a more stable forecasting model. In the empirical part, we choose to split the sample into 10 subsamples.

Notably, one might argue that the analysis here might be problematic since the time series structure could be destroyed during the cross-validation. However, the cross-validation here is not simply resampling over the original sample, instead block sampling methods are employed to keep the time series structure. Moreover, to have a more fair comparison, this paper does not expand

the model into ARIMA or other popular time series models, since the introduction of lags might offset the true effects from the queries. The extension could be made in the future, and model (1) can be extended to any desired form, while this paper only contributes a benchmark.

Data collection

We collect the daily tourist records from a five-star resort, Jiuzhaigou, in Sichuan province in China. The reason why we choose this resort is that this resort is a perfect representative of those resorts attracting considerable amount of tourists everyday, and it has stably long-run inputs than any other new started resorts. In addition, compared to other resorts, the data it has is well-recorded, thus the data quality can be guaranteed.

The corresponding search query data is collected from a search engine website, provided by a dominant search engine service supplier, Baidu, in China. This company takes huge market shares so that the sample selection bias, on an aggregated level, is mitigated than any other data sources. We collect the data using an enumerate manner, namely we first pick several representative queries by our subjective knowledge and then take the recommended queries from the recommendation system of the search engine. Once either few new queries are recommended or the new words are totally irrelevant from our subjective knowledge, we stop iterating.

Finally, we have 148 relevant queries as candidates as indicators for forecasting. The queries collected reflect numerous demands on the tours to Jiuzhaigou, as summarized in the following categories: the name of the resort and its relevant information, the detailed information on the scenes inside the resort, the travelling information regarding the plans, suggestions, routes and other resorts nearby etc. Since the queries are in Chinese, in which the forming of words are quite different from English, we code each of them into an English category plus a number as an identifier. For example, the fifth query about travel information is coded as "TravellInformation5". It can be seen that some queries are intuitively important since most of the travellers will search them while some queries recommended by the website are not a must for travellers to search. Subjective selection of queries become invalid given huge amount of candidate queries.

Therefore, it is needed to employ variable selection models to select the proper queries. Currently, we collect both the actual tourists records and the search query data from the 1st of June, 2012 to the 31th of August, 2014, with 822 observations on a daily basis.

Empirical comparisons

This section shows the comparisons of forecasting performance and variable selection between BMA and

Lasso. For the BMA models, the forecasting performance evaluation is based on the posterior means of the predicted tourist amounts in the out-of-sample evaluation. For the Lasso, we apply 10-folded cross-validation method, and use the best model to do out-of-sample evaluation. For the stability of the results, the evaluation of each model takes the average of 5 times evaluations. Compared to Lasso, some details of BMA are given, because the basic settings of BMA are much flexible than Lasso which only has one tuning parameter λ . Even though the Lasso is of simpler form, we also estimate Ridge estimator as a contrast.

To conduct this, we will consider the effects of different model priors, different MCMC sampling methods and different Zellner's g s. Since the Zellner's g acts as a shrinkage element which is viewed as the most important feature of BMA models, in the following empirical parts we first compare the models with different model priors and MCMC methods, then choose the best model featuring different Zellner's g to compare with Lasso.

Therefore after this short introduction, we then discuss the forecasting performance and the variable selection, and some estimate results of both BMA and Lasso are given prior to the comparison.

Model estimates

Before conducting the comparison, we first give some model estimates to have some straightforward ideas about the models involved, including the posterior inclusion probabilities of BMA models with different model priors, the quality of different MCMC methods and the shrinkage results of Lasso compared with Ridge estimator.

BMA estimates

We first estimate the BMA models. Since the performance model priors might be sensitive to the model's prior, that is $p(M_\gamma)$, we employ three different model priors to be compared for robustness, that is, uniform prior, binomial prior, and the beta-binomial prior. In addition, to avoid the case where certain MCMC sample is not effective in this particular research, we employ both the birth-death sampler and the reversible-jump sampler (Madigan et al., 1995).

In simple words, the birth-death sampler, in each round, randomly pick a variable to drop or to add, depending on whether the chosen variable is in the current model or not. The reversible-jump sampler assigns birth-death method a probability of 50% and for the other 50%, it swaps one of the variables already in the model with another one which is not in the current model. Importantly, the posterior mean of each parameter estimate is the weighted average across the top 200 models. To be

noticed, if a variable is not included in a certain model, the coefficient, that is, zero, is also accounted into this average.

In Figure 1, "uniform" means the model prior is of the uniform type, "fixed" means that the prior is of the binomial type with expectation of 20, and "random" means that the prior is of beta-binomial type. We can see that the model with uniform prior shrinkages smoothly and model with binomial prior shows unstable performance when assigning inclusive probabilities to variables. Nevertheless, all the priors make no huge distinction with each other (Figure 2).

The MCMC methods might suffer from some low-quality approximations, and one way to examine them is to compare them with the analytical results, which are those smooth and red lines in Figure 2. We find that the reversible jump MCMC has better quality than birth-death one. In general, the models with Beta-binomial are accompanied with the worst MCMC approximation, and the models with uniform, that is, a non-informative prior, are poorly approximated through MCMC because the uniform prior requires all the models being assigned with equal weights.

Lasso estimates

The algorithm used to implement Lasso is the coordinate descent method. We first select the penalty parameter, λ without cross-validation. The number of λ sequence employed is 100. This is the same to ridge regression. Figure 3 explicitly shows the hard-thresholding property of Lasso. In this empirical question, we have 148 independent variables in total. The horizontal axis is the penalty term realization of Equation (10) on the left panel in Figure 3, and the penalty term realization of Equation (11) on the right panel. As we can find, the shrinkage in Lasso is stricter than in Ridge. As the penalty becomes larger, given λ , less variables will be screened out in Lasso; while in ridge regression, the shrinkage is milder. Obviously, the chosen model given λ might be unstable somehow, therefore when we conduct forecasting we use cross-validation to select proper models, and the criterion of model selection of the cross-validation consists of three types: Deviance, MAE (Mean Absolute Error), and MSE (Mean Squared Error). The results will be discussed in detail in the forecasting performance section.

Forecasting performance

Now we make use of the estimates, and employ the test sample to calculate the RMSE (Rooted Mean of Squared Errors) to make comparison. We use 800 out of 822 observations to estimate and select the model, and let the other 22 to be an out-of-sample for evaluation.

The outcome can be read from Table 1, and the overall

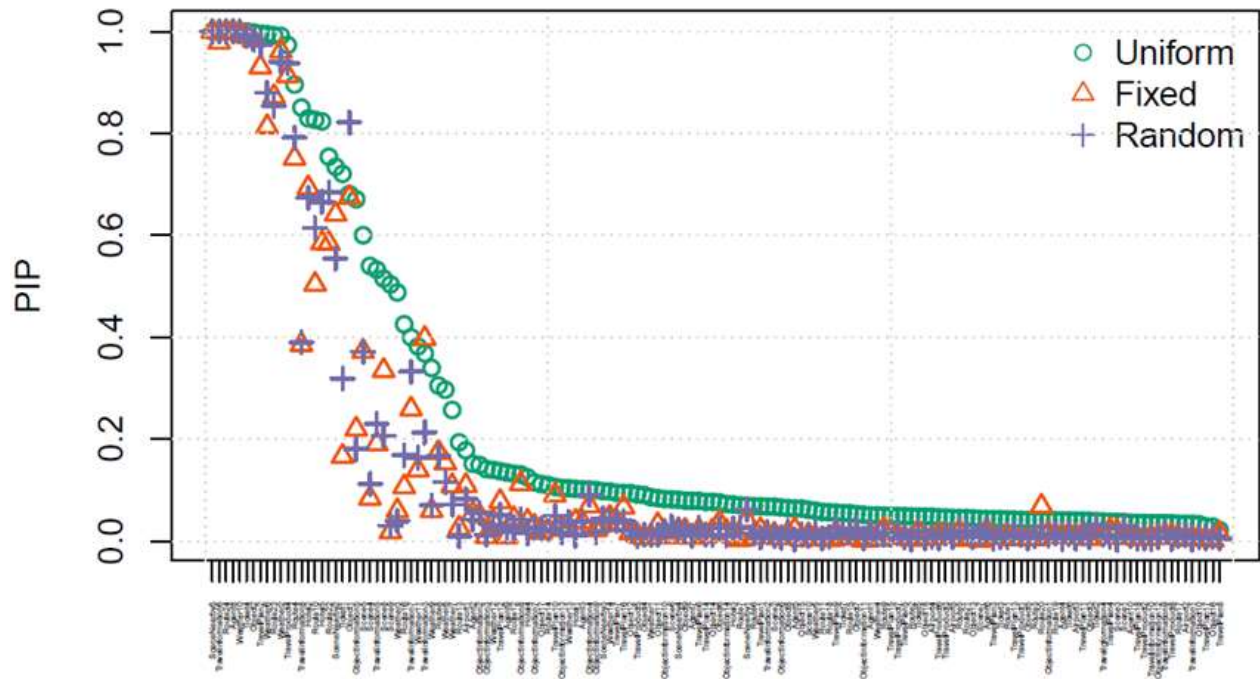


Figure 1. Posterior inclusive probabilities with different model priors.

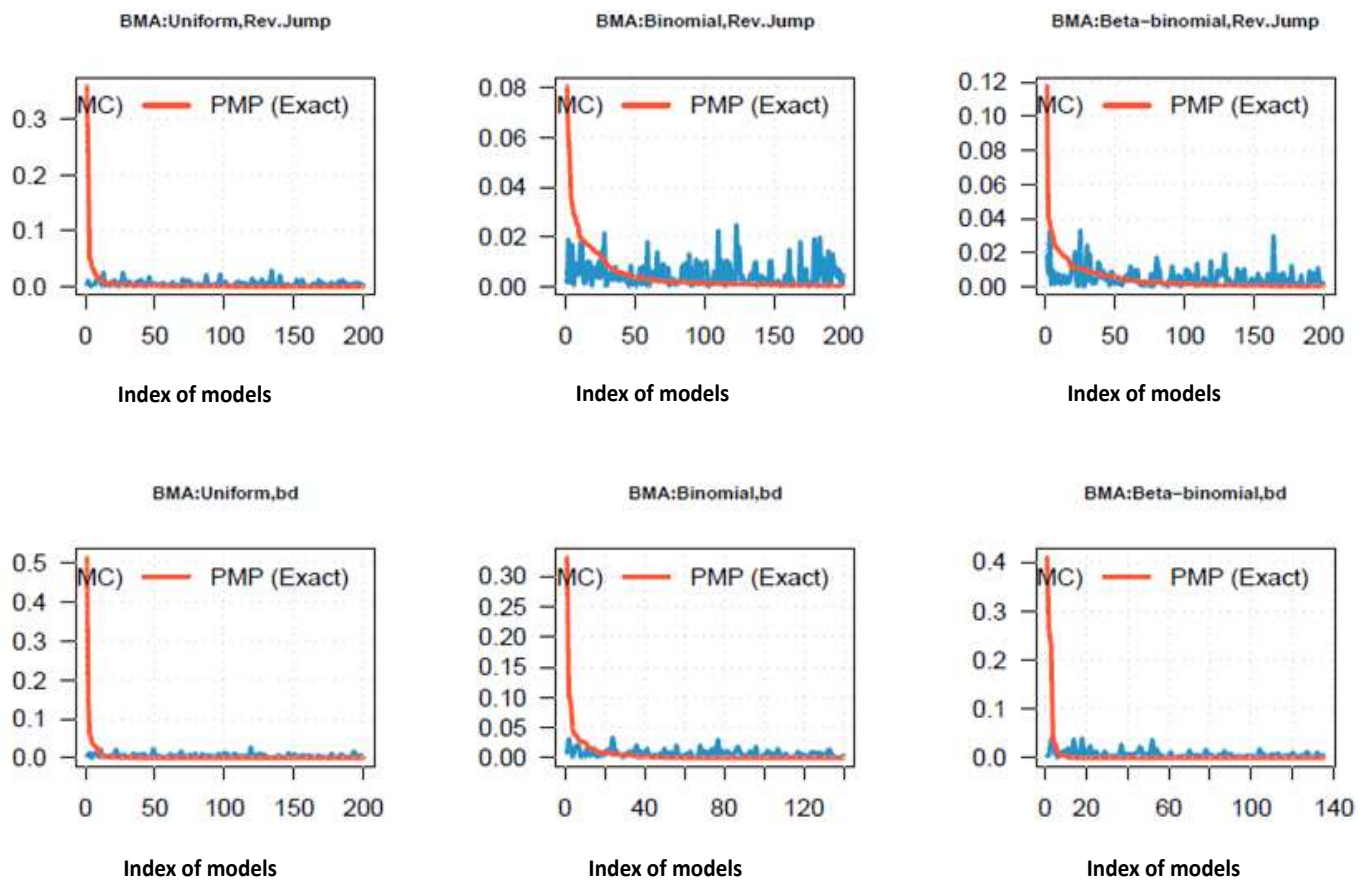


Figure 2. Posterior model probabilities of MCMC and analytical results.

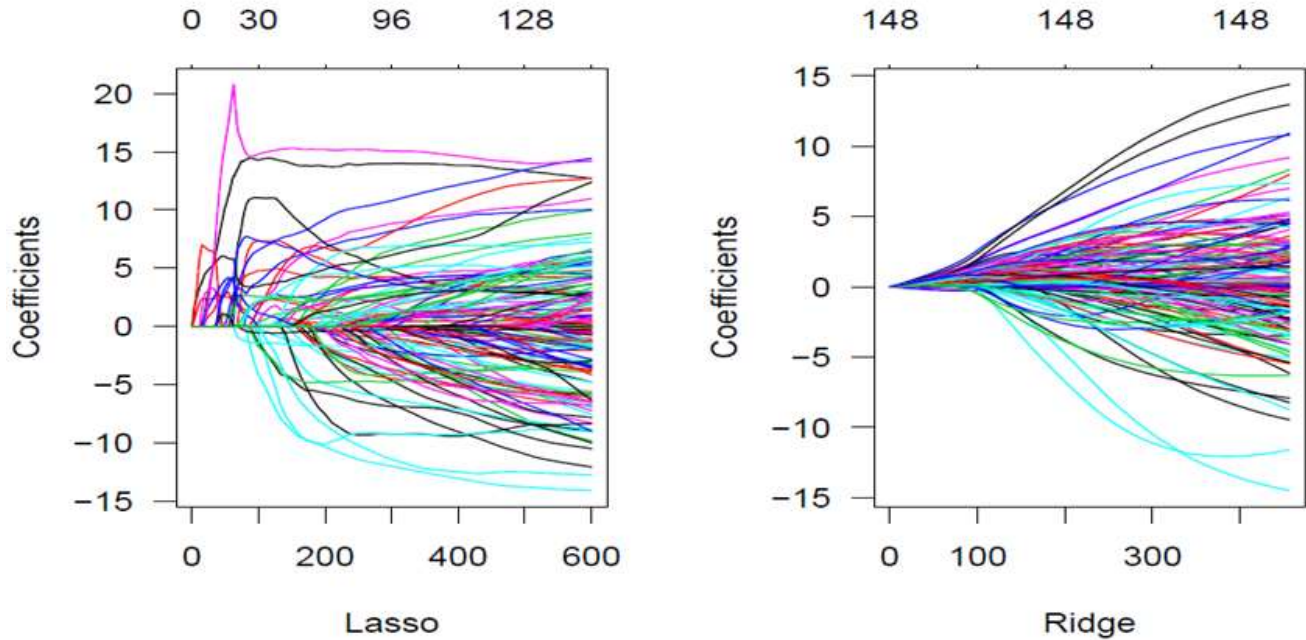


Figure 3. Model selection of lasso and ridge regression.

Table 1. Forecasting performance.

Model	Detail	RM SE
BMA	Uniform, Birth-death	3892.62
	Binomial, Birth-death	3903.516
	Beta-binomial, Birth-death	3912.273
	Uniform, Reversible Jump	3644.754
	Binomial, Reversible Jump	3861.538
	Beta-binomial, Reversible Jump	4075.911
Lasso	Lasso	5214.74
	Ridge	7978.638
	Lasso-CV-deviance	3934.523
	Lasso-CV-mae	4014.339
	Lasso-CV-mse	3794.487
BMA	Uniform, UIP	3517.69
(Zellner's g)	Uniform, EBL	3571.29
	Uniform, Hyper-g	4044.663

distribution of residuals can be seen in Figure 4. It can be read that the best performance is from the BMA with uniform prior and the reversible jump MCMC method. In general, we find that the forecasts using uniform priors are the best, using binomial ranks second and beta-binomial performs worst among the BMA forecasts. This might be an evidence that the non-informative priors actually are

good at forecasting. For the lasso, the forecast with MSE cross-validation performs best, and this might be due to that the criterion we use for out-of-sample evaluation is RMSE, which is the closest criterion to MSE. From the numerical results, we do not find whether BMA outperform Lasso, or the other way around in general. If we further look at Figure 4, we also find that the residual sequences

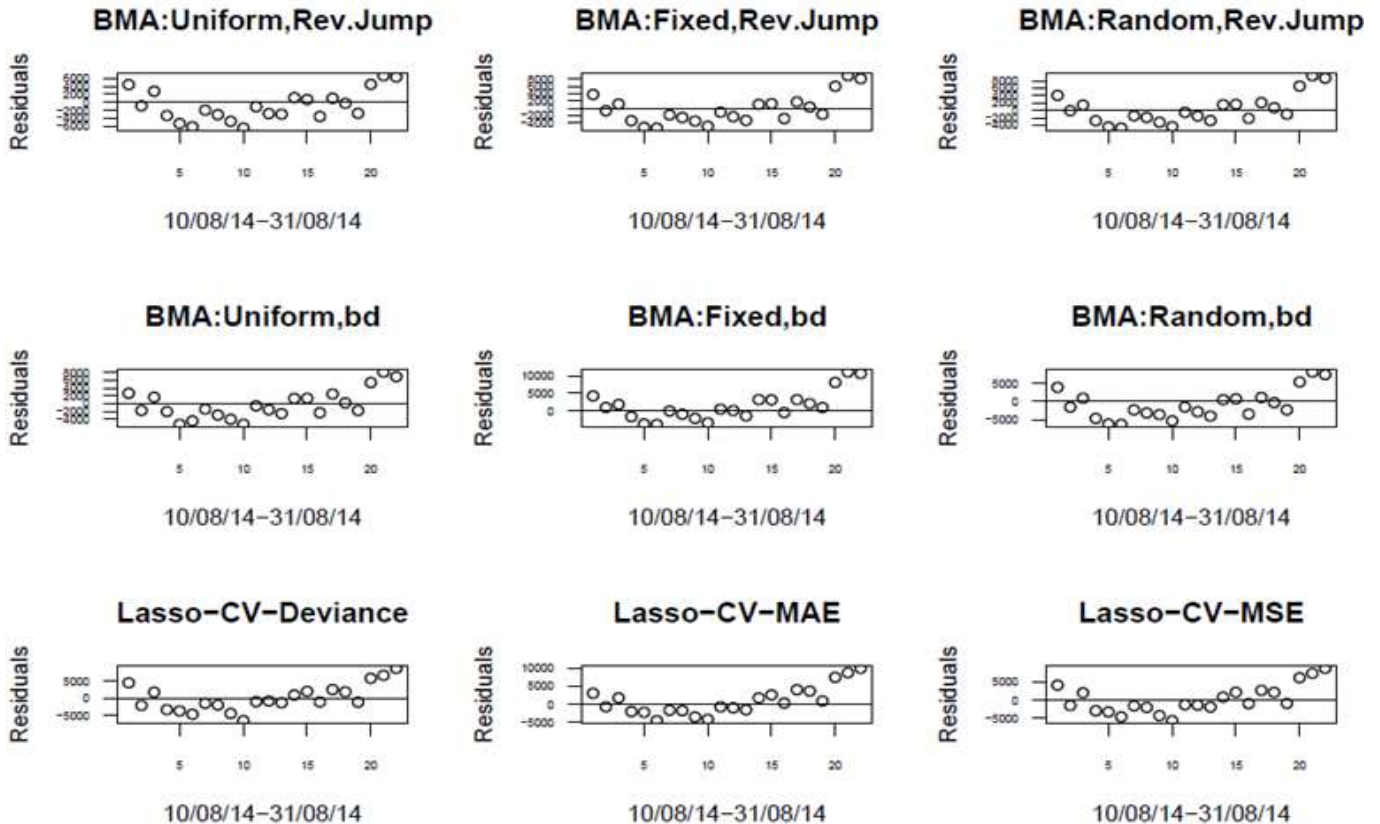


Figure 4. Residual plots for BMA and lasso.

have similar patterns, that is, some fluctuations or seasonal patterns on a weekly basis. This implies that the models can be improved considering more complex time series variants such as ARIMA, ARCH, etc. However, in this paper, the linear model is sufficient for comparison of query selections so we will not complicate the settings.

Till now, we only take various model priors and MCMC methods into account. As a crucial element, the Zellner's g seems to play more important roles in forecasting and model selection. We choose three types of g s in our research: fixed with $g = N$, the Empirical Bayes, Local type (Hansen and Yu (2001)), and the hyper- g type (Liang et al., 2008).

Based on the "best" model, that is, model with prior uniform and with reversible jump sampler, we conduct the evaluation 5 times, and take the average. The RMSE are 3517.69, 3571.29 and 4044.66, respectively. It can be read that, still, the simple rule wins, since the model with uniform model prior and fixed g has the best performance. It outperforms Lasso, but we cannot say BMA is definitely better than Lasso, since the performance difference is not huge. Nevertheless, the performance reached by both BMA and Lasso is satisfying, compared to the previous results, even though there is space to improve (Figure 4).

Model size and variable selection

First of all, the variable selection here is somewhat confusing in BMA context, since the philosophy of model averaging is proposed against model selection. However, we can treat the posterior inclusion probability (PIP) as the measure of likelihood that certain variables should be included or not. The PIPs demonstrate the posterior probabilities that a variable is included in the models. This is calculated through the summation of the posterior probability of those best models including this variable.

It would be trivial to compare the variable selection if the model sizes are not taken into consideration. In other words, the larger model has more variables, and it has high probability to contain all the variables selected by the small model. Therefore, we first compare the model sizes across different models.

For the BMA models, the model priors affect the model sizes, however, since the likelihoods dominate the posterior as sample size becomes larger, the model sizes are mainly affected by the information underlying the data. In Figure 5, we compare the models with different model priors and Zellner's g s. For the Uniform prior, the expected model size is around 70, while the posterior size is around 30. For the Binomial prior, since we set the

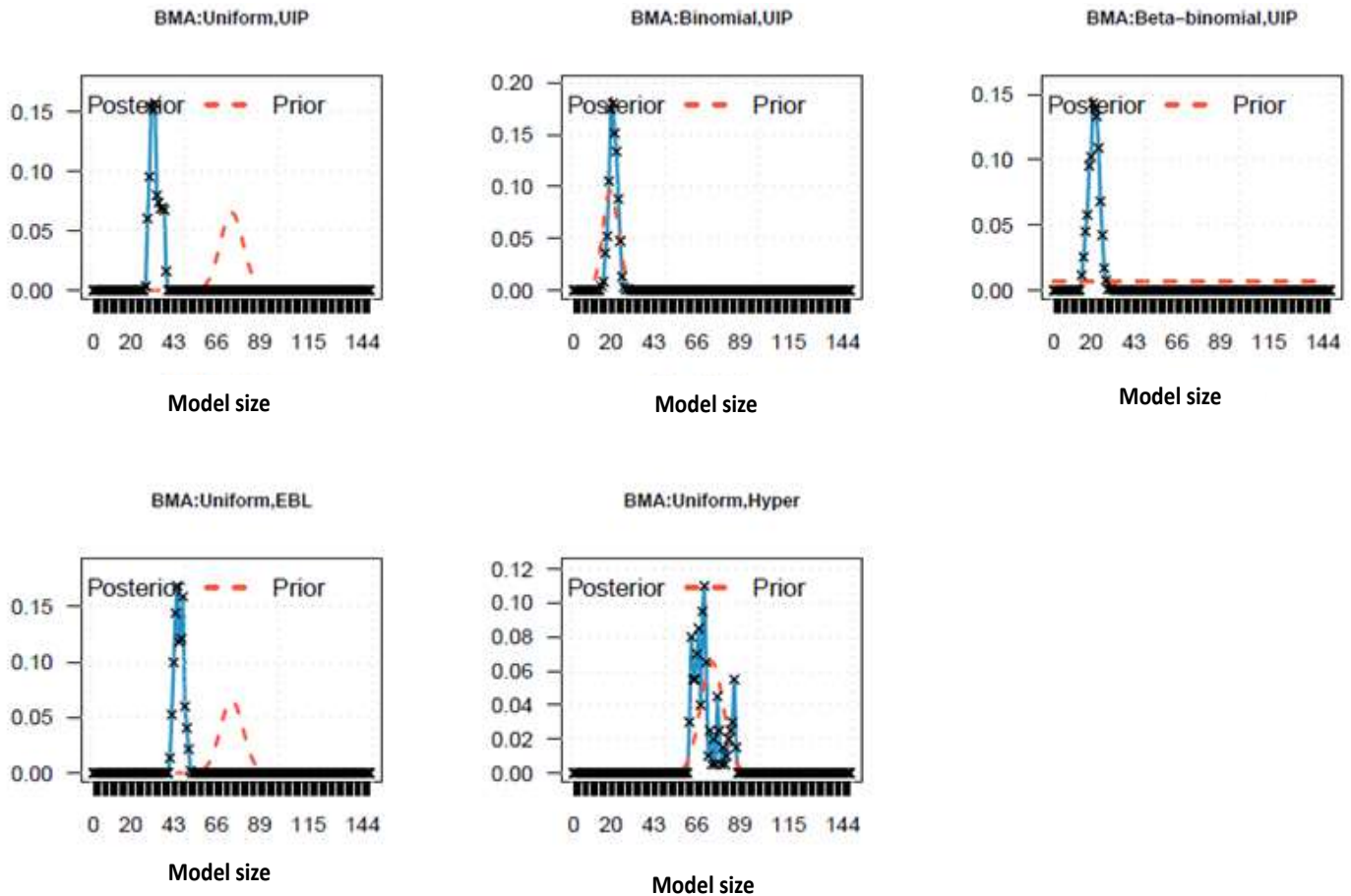


Figure 5. Model size of BMA.

expected model size as 20, the posterior is also around 20. For the Beta-binomial, the prior is totally flat over the horizontal axis, while the posterior is still around 20. Zellner's g does affect the posterior model sizes, and we find that more complex g yields more complex posterior size distributions. In addition, all the models with uniform priors has larger model sizes because the prior "drags" the posterior toward it (Figure 5). The model sizes of Lasso with cross-validation, as in Figure 6, are around 30 or slightly above. Cross-validation with deviance selects the best model with the size around 30, similar to the mae type. However, the MSE type are slightly larger.

All the comparisons earlier mentioned show that the model sizes differ little between Lasso and BMA. This implies that the intrinsic model size supported by the data is around 30 to have a better forecasting performance. Therefore, it is interesting to see whether Lasso and BMA select similar variables or not. If they select several common variables, the practitioners can then pay more attention to those common queries.

The variable selection results are given in Table 2, and the variables on the list are selected by the Lasso with deviance. If other methods have extra selected variables,

then we currently do not report them in the table, since those variables are not important in magnitude, and this does not affect our results because we can see that the estimates of those on-the-list variables are similar. For the BMA estimates, we not only show the parameter estimates, but also give the posterior inclusive probabilities, which is an indicator whether certain variables are not important from BMA point of view. Of course, all the candidate variables has been assigned a PIP, but most of them are very low (Figure 6).

Conclusions

Forecasting tourist arrivals is beneficial. Using online information to conduct forecasts is of great value. In the past, we have no choices; while nowadays, we have too many choices. Selecting proper queries to be built in the forecasting models plays crucial roles among practitioners and academia. Previous research reveals the limitations of conducting simple index composition, or variable selection based on correlation coefficient or other criteria.

Lasso and BMA are both ways to select statistically

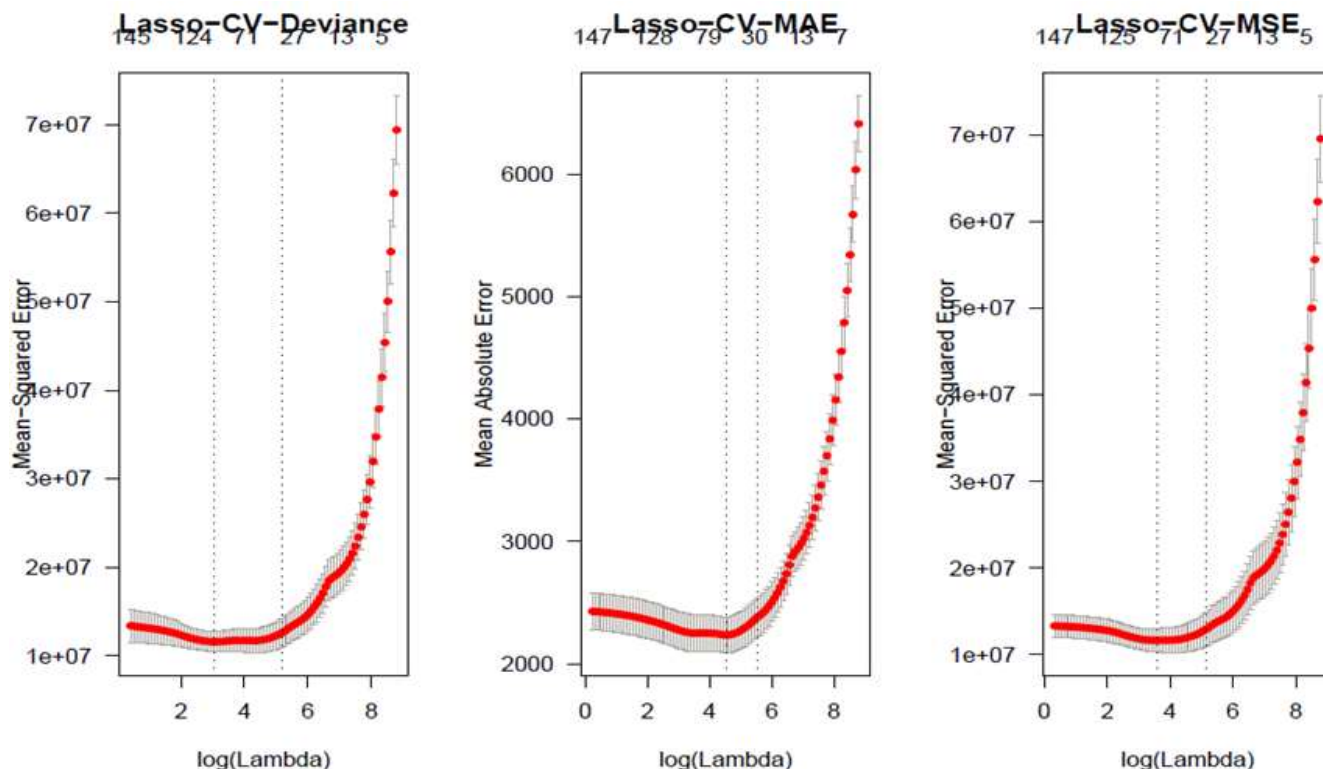


Figure 6. Model size of Lasso.

Table 2. Variable selection results.

Variable	Deviance	Lasso			BMA	
		MAE	MSE	UIP	EBL	Hyper
Object1	0.07	0.07	0.06	0.16 (0.80)	0.05 (0.30)	0.16 (0.74)
Object2	4.60	4.78	4.76	5.06 (1.00)	4.69 (1.00)	3.84 (1.00)
Object8	1.55	1.91	2.13	0.26 (0.09)	0.58 (0.16)	2.09 (0.68)
ObjectInformation 5	-2.73	-3.99	-8.35	-6.13 (0.49)	-9.13 (0.71)	-16.89 (1.00)
ObjectInformation 8	7.20	7.34	7.13	0.32 (0.06)	1.12 (0.24)	4.50 (1.00)
Scene2	3.53	3.77	5.70	2.73 (0.34)	1.56 (0.19)	9.26 (1.00)
Scene4	0.43	1.52	4.93	6.42 (0.49)	11.77 (0.84)	15.72 (1.00)
Scene5	14.4	14.2	14.17	5.75 (0.41)	13.01 (0.95)	11.16 (1.00)
Scene8	6.24	6.53	7.72	3.81 (0.32)	3.86 (0.37)	10.26 (1.00)
SceneNearby2	0.004	0.04	0.21	0.27 (0.51)	0.57 (0.98)	0.18 (0.39)
TravelPlan3	0.04	-0.18	1.4	0.07 (0.03)	0.04 (0.03)	1.31 (0.36)
TravelPlan10	11.03	11.06	10.12	0.53 (0.10)	0.09 (0.03)	0.59 (0.21)
Route10	-0.45	-1.34	-4.23	-12.71 (1.00)	-12.44 (1.00)	-11.59 (1.00)
Route12	-0.73	-1.69	-4.09	-0.76 (0.15)	-1.11 (0.16)	-5.56 (1.00)
Route23	0.49	1.37	5.42	14.06 (0.98)	14.96 (1.00)	13.55 (1.00)
Agent1	1.78	2.04	2.62	0.71 (0.14)	0.47 (0.14)	3.47 (0.75)
Agent8	-1.38	-1.41	-1.54	-1.81 (1.00)	-1.92 (1.00)	-2.17 (1.00)
Agent10	-0.51	-0.53	-0.53	-0.03 (0.12)	0.00 (0.08)	-0.10 (0.82)
Weather2	2.60	2.55	2.50	3.58 (1.00)	3.94 (1.00)	4.21 (1.00)
Weather4	0.10	0.17	0.14	-0.03 (0.06)	0.00 (0.00)	0.00 (0.00)
Weather6	3.37	3.45	3.87	8.03 (1.00)	9.13 (1.00)	10.17 (1.00)
Weather9	0.63	1	-	-0.96 (0.15)	-3.30 (0.50)	-3.37 (0.58)

Table 2. Contd.

Ticket2	1.48	1.16	0.40	0.07 (0.06)	1.01 (0.24)	0.94 (0.70)
Ticket4	14.62	14.8	15.23	15.13 (1.00)	14.51 (1.00)	13.47 (1.00)
Hotel1	1.96	2.17	2.35	1.26 (0.17)	0.39 (0.10)	6.14 (0.72)
Hotel2	7.51	7.37	6.82	6.48 (0.65)	10.65 (0.97)	6.62 (0.51)
Travel product4	1.80	2.19	3.38	7.75 (1.00)	5.66 (0.79)	0.00 (0.00)
Other variable	No	No	Yes	Yes	Yes	Yes
RMSE	3934.523	4014.339	3794.487	3517.687	3571.288	4044.663

important variables. We employ these methods to conduct forecasts, and the models exhibit satisfying performance. In addition, the difference of performance between two approaches is not huge, while BMA with uniform model priors and unit information prior on Zellner's g performs best.

Lasso demonstrates strict shrinkage and selects fewer queries entering the models, while BMA selects queries based on posterior inclusive probabilities. There are several important queries which are both selected by Lasso and BMA, while some other variables chosen by BMA are screened out by Lasso, vice versa.

It is still open for expanding the models featured with more complex components, however computational issue is another vital point to be studied in depth, especially with more queries implemented by MCMC methods.

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

REFERENCES

- Bangwayo-Skeete PF, Skeete RW (2015). Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tour. Manage.* 46:454-464. <http://dx.doi.org/10.1016/j.tourman.2014.07.014>
- Brynjolfsson E, Geva T, Reichman S (2016). Crowd-Squared: Amplifying the Predictive Power of Search Trend Data. *Management Information Systems Q.* 40(4):941-961. <http://dx.doi.org/10.2139/ssrn.2513559>
- Butler D (2013). When Google got flu wrong. *Nature* 494(7436):155. <http://dx.doi.org/10.1038/494155a>
- Doppelhofer G, Miller RI (2004). Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach. *Am. Econ. Rev.* 94(4):813-835. <https://doi.org/10.1257/0002828042002570>
- Feldkircher M (2012). Forecast combination and Bayesian model averaging: A prior sensitivity analysis. *J. Forecasting* 31(4):361-376. <http://dx.doi.org/10.1002/for.1228>
- Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232):1012-1014. <http://dx.doi.org/10.1038/nature07634>
- Hansen MH, Yu B (2001). Model selection and the principle of minimum description length. *J. Am. Statist. Assoc.* 96(454):746-774. <http://dx.doi.org/10.1198/016214501753168398>
- Hoeting JA, Madigan D, Raftery AE, Volinsky CT (1999). Bayesian model averaging: a tutorial. *Stat. Sci.* pp. 382-401. <http://dx.doi.org/10.1214/ss/1009212519>
- Ley E, Steel MF (2009). On the effect of prior assumptions in Bayesian model averaging with applications to growth regression. *J. Appl. Econometrics*, 24(4):651-674. <http://dx.doi.org/10.1002/jae.1057>
- Li X, Pan B, Law R, Huang X (2017). Forecasting tourism demand with composite search index. *Tour. Manage.* 59:57-66. <http://dx.doi.org/10.1016/j.tourman.2016.07.005>
- Liang F, Paulo R, Molina G, Clyde MA, Berger JO (2008). Mixtures of g priors for Bayesian variable selection. *J. Am. Statist. Assoc.* 103(481):410-423. <http://dx.doi.org/10.1198/016214507000001337>
- Madigan D, York J, Allard D (1995). Bayesian graphical models for discrete data. *International Statistical Review/Revue Internationale de Statistique*, pp. 215-232. <http://dx.doi.org/10.2307/1403615>
- Moral-Benito E (2015). Model averaging in economics: An overview. *J. Econ. Surv.* 29(1):46-75. <http://dx.doi.org/10.1111/joes.12044>
- Scott SL, Varian HR (2014). Bayesian variable selection for nowcasting economic time series. In *Economic Analysis of the Digital Economy* University of Chicago Press pp. 119-135. <http://dx.doi.org/10.7208/chicago/9780226206981.003.0004>
- Tibshirani R (2011). Regression shrinkage and selection via the lasso: A retrospective. *J. Royal Stat. Soc.* 73(3):273-282. <http://dx.doi.org/10.1111/j.1467-9868.2011.00771.x>
- Yang X, Pan B, Evans JA, Lv B (2015). Forecasting Chinese tourist volume with search engine data. *Tour. Manage.* 46:386-397. <http://dx.doi.org/10.1016/j.tourman.2014.07.019>