Ann Inst Stat Math (2010) 62:109–116
DOI 10.1007/s10463-009-0250-4

Does a Bayesian approach generate robust forecasts? Evidence from applications in portfolio investment decisions

Chih-Ling Tsai · Hansheng Wang · Ning Zhu

Received: 26 April 2008 / Revised: 17 December 2008 / Published online: 10 July 2009

Abstract We employ a statistical criterion (out-of-sample hit rate) and a financial market measure (portfolio performance) to compare the forecasting accuracy of three model selection approaches: Bayesian information criterion (BIC), model averaging, and model mixing. While the more recent approaches of model averaging and model mixing surpass the Bayesian information criterion in their out-of-sample hit rates, the predicted portfolios from these new approaches do not significantly outperform the portfolio obtained via the BIC subset selection method.

Keywords Model selection · BIC · Model averaging · Model mixing · Stock predictability · Financial markets

1 Introduction

In the context of financial economics, forecasting future stock market movement has always been a focal topic. Over a century ago, the famous Austrian economist William S. Jevons related stock market returns to sun spot activities (Peart 1996). With the development of statistics and computational science, practitioners and academics

C.-L. Tsai (✉) · N. Zhu
Graduate School of Management, University of California-Davis,
Davis, CA 95616-8609, USA
e-mail: cltucd@gmail.com

N. Zhu
e-mail: nzhu@ucdavis.edu

H. Wang
Department of Business Statistics and Econometrics, Guanghua School of Management,
Peking University, 100871 Beijing, People's Republic of China
e-mail: hansheng@gsm.pku.edu.cn

in finance have been striving to predict future stock returns over the past few decades (e.g., see Malkiel 2004). As it is arguably more difficult to predict the future movement of an individual stock price, numerous studies are devoted to understanding whether one can successfully predict the ups and downs of the stock market (i.e. indices of different stock markets). Please refer to Rachev et al. (2006) for a comprehensive list of references.

Based on extant studies, there are various findings and conclusions regarding stock market predictability. Many researchers claim that some variables can reliably predict future stock returns, at least within the selected sample period and the given estimation methods (see Campbell 1987; Ang and Bekaert 2007). Furthermore, Chen et al. (1986) and Bossaerts and Hillion (1999) argue that macro-economic variables, risk spread, and dividend yield are useful in predicting future returns. For a nice summary regarding the variables related to stock market prediction, we refer to Cremers (2002). This immediately raises a question of whether all those variables are relevant to stock market predictions. Moreover, many studies acknowledge that their observed predictability is robust only within the sample period and not particularly useful in making out-of-sample predictions. Other notable critics question predictability and attribute the successes in existing studies to data-snooping and model uncertainty (Black 1993; Lo and MacKinlay 1997).

One way to partly address the data snooping problem is via the Bayesian approach. Under the Bayesian framework, although the initial choice of variable (prior) still suffers from data snooping and a particular specification, the problem is minimized in the updated choice from the posterior exploration (Leamer 1978; George and McCulloch 1993). As demonstrated in the literature, when the sample size is sufficiently large, the posterior of a given candidate model depends more on its Bayesian information criterion (BIC) and far less on the prior (Schwarz 1978; Avramov 2002; Cremers 2002). Thus, the optimal model can be selected by the best subset selection method, BIC. A similar approach can also be implemented for the renowned Akaike information criterion (Akaike 1973, AIC).

In theory, the BIC-type best subset selection method is optimal if there indeed exists a "true" model and the sample size is sufficiently large. However, many researchers argue that using only the best subset selection method (including BIC), which means putting all weight in one selected "optimal model", is too risky. Thus, it is better to combine a number of competitive models with appropriate weights (e.g., see Akaike 1978a; Buckland et al. 1997; Hoeting et al. 1999; Burnham and Anderson 2002; Yang 2001, 2003). Such an innovative idea generates two classes of model combination methods, namely, model averaging and model mixing.

The method of model averaging is very simple. For example, in the context of stock market prediction, one first obtains forecasting results based on a number of competing models and then averages those prediction results with appropriate weights. Determining the weights is a very important and challenging task. Various weighting schemes have been developed in the literature (see Hjort and Cleaskens 2003 for a good summary and discussion). Among the various model averaging methods, the Bayesian model averaging (BMA) approach of Hoeting et al. (1999) is commonly used due to its mathematical simplicity and practical usefulness. Simply speaking, the BMA method assigns a weight to the candidate model according to its posterior probability. For a given candidate model $\mathcal{S}$, one is able to show that such a posterior probability is proportional to $\exp(-\text{BIC}_{\mathcal{S}}/2)$ as long as the sample size is sufficiently large (Hastie et al. 2001), where $\text{BIC}_{\mathcal{S}}$ stands for the BIC score associated with the model $\mathcal{S}$.

The BMA method represents a large class of model averaging methods that assign the model weights according to some pre-specified in-sample selection criterion. For example, the BMA method assigns the weight according to BIC, while the frequentist model averaging method of Hjort and Cleaskens (2003) assigns the weight according to AIC (See also Akaike 1978b, 1979 for earlier discussions). If prediction accuracy is our ultimate concern (e.g., the stock market prediction discussed here), why can not one assign the model weight according to its prediction accuracy? Motivated by this intuition, Yang (2001, 2003) proposed the method of model mixing. In essence, the mixing method first splits the data into two subsamples of equal size. Next, it uses the former subsample to build up the model and the later subsample to evaluate the model's forecasting accuracy with the appropriate criterion. Finally, different combination weights are assigned to the competing models according to their estimated forecasting accuracy.

While the theoretical properties of best subset selection, model averaging, and model mixing have been extensively investigated in the literature, empirical studies in the field of finance are relatively limited. This motivates us to apply the above three approaches in the realm of financial economics. Based on the US stock market data obtained from the Center for Research in Security Prices (CRSP) database, we find that these approaches lead to a better prediction than a purely random process. Furthermore, averaging and mixing yield a slightly higher hit rate in predicting next-period stock market returns when compared to BIC. However, none of the methods examined provide a significant improvement over BIC in the prediction of portfolio performance. This important finding indicates that the forecast obtained from BIC is robust for the purpose of portfolio investment decisions.

The remainder of this paper is organized as follows. Section 2 outlines the three forecasting methods and describes the data used in the study. Section 3 presents our main findings and discusses the results before we conclude in Sect. 4.

2 The method and data

2.1 Methodology

Let (X_i, Y_i) be the observation collected from the ith subject, where $Y_i \in \mathcal{R}^1$ is the response of interest and $X_i = (X_{i1}, \ldots, X_{ip})^\top \in \mathcal{R}^p$ is the associated predictor for $i = 1, \ldots, n$. Furthermore, we use generic notation $\mathcal{S} = \{j_1, \ldots, j_d\}$ to denote a candidate model, which includes $X_i^{(\mathcal{S})} = (X_{i,j_1}, \ldots, X_{i,j_d})^\top$ as relevant predictors. For a given $\mathcal{S}$, the working linear model can be constructed as

$$Y_i = X_i^{(\mathcal{S})} \beta^{(\mathcal{S})} + \varepsilon_i, \tag{1}$$

where $\beta^{(\mathcal{S})} = (\beta_{j_1}, \ldots, \beta_{j_d})^\top \in \mathcal{R}^d$ is the subvector of $\beta = (\beta_1, \ldots, \beta_p)^\top \in \mathcal{R}^p$. For the sake of prediction, we consider (X_0, Y_0) to be an independent observation, where $X_0 = (X_{01}, \ldots, X_{0p})^\top \in \mathcal{R}^p$. We then assess how to predict the value of Y_0 based on X_0 via the following three prediction methods.

2.1.1 The BIC method

This is a very typical method, which predicts the future observation according to the best selected candidate model. More specifically, for each candidate model $\mathcal{S}$, we compute its ordinary least squares estimator (denoted by $\hat{\beta}_{\mathcal{S}}$) and then calculate its associated BIC criterion,

$$\mathrm{BIC}_{\mathcal{S}} = n \log \left\{ \sum_{i=1}^{n} \left(Y_i - X_i^{(\mathcal{S})} \hat{\beta}_{\mathcal{S}} \right)^2 \right\} + |\mathcal{S}| \times \log n, \tag{2}$$

where $|\mathcal{S}|$ denotes the number of predictors included in $\mathcal{S}$. We subsequently find the best model, $\mathcal{S}_{\mathrm{BIC}} = \mathrm{argmin}_{\mathcal{S} \subset \mathcal{S}_F} \mathrm{BIC}_{\mathcal{S}}$, where $\mathcal{S}_F = \{1, \ldots, p\}$ stands for the full model. Accordingly, we predict Y_0 by $\hat{Y}_0^{\mathrm{BIC}} = X_0^{(\mathcal{S}_{\mathrm{BIC}})\top} \hat{\beta}_{\mathcal{S}_{\mathrm{BIC}}}$.

2.1.2 Model averaging

In contrast to applying the "best" model selected by BIC to make forecasts, we are able to employ the BMA method to predict future observations by averaging a number of competitive models. In this article, we consider a total number of 2^p competitive models (including the null model). For each candidate model $\mathcal{S}$, we compute its OLS estimator $\hat{\beta}_{\mathcal{S}}$ with the corresponding weight $w_{\mathcal{S}} = \exp(-\mathrm{BIC}_{\mathcal{S}}/2)$. Then the resulting average estimator is

$$\hat{\beta}_{\mathrm{BMA}} = \left(\sum_{\mathcal{S}} w_{\mathcal{S}} \hat{\beta}_{\mathcal{S}} \right) \left(\sum_{\mathcal{S}} w_{\mathcal{S}} \right)^{-1}. \tag{3}$$

As a result, we predict Y_0 by $\hat{Y}_0^{\mathrm{BMA}} = X_0^\top \hat{\beta}_{\mathrm{BMA}}$.

2.1.3 Model mixing

Analogous to BMA, mixing is also a model combination method but with a different weighting scheme. More specifically, we first split the data into $\mathcal{D}_1 = \{X_i, Y_i\}_{i=1}^{n/2}$ and $\mathcal{D}_2 = \{X_i, Y_i\}_{i=n/2+1}^{n}$. For a given candidate model, we next compute its OLS estimator, $\tilde{\beta}_{\mathcal{S}}$, based on dataset $\mathcal{D}_1$. Furthermore, the variance of the residual is estimated by $\tilde{\sigma}_{\mathcal{S}}^2 = (n/2)^{-1} \sum_{i=1}^{n/2} (Y_i - X_i^{(\mathcal{S})} \tilde{\beta}_{\mathcal{S}})^2$. Subsequently, we evaluate the model's prediction error by $D_{\mathcal{S}} = (n/2)^{-1} \sum_{i=n/2+1}^{n} (Y_i - X_i^{(\mathcal{S})} \tilde{\beta}_{\mathcal{S}})^2$. Then, the mixing estimator is given as follows:

$$\hat{\beta}_{\text{MIX}} = \left(\sum_{\mathcal{S}} \tilde{w}_{\mathcal{S}} \hat{\beta}_{\mathcal{S}}\right)\left(\sum_{\mathcal{S}} \tilde{w}_{\mathcal{S}}\right)^{-1}, \tag{4}$$

where $\tilde{w}_{\mathcal{S}} = \tilde{\sigma}_{\mathcal{S}}^{-n/2} \exp(-\tilde{\sigma}_{\mathcal{S}}^{-2} D_{\mathcal{S}}/2)$. Finally, we predict Y_0 by $\hat{Y}_0^{\text{MIX}} = X_0^{\top} \hat{\beta}_{\text{MIX}}$.

2.2 The dataset

Following the extant literature on stock market predictability, we use monthly data on the publicly available economic variables to forecast the return of U.S. stock market in the subsequent month. Consistent with the majority of existing studies, we employ the value-weighted CRSP stock return index as the dependent variable of future stock returns. In addition, we consider 12 predictive variables that have been shown to be important in forecasting future stock returns (see Cremers 2002). These independent variables are: S&P 500 dividend yield (the total standard&poor 500 dividend payout divided by standard&poor 500 price level), S&P 500 earnings/price ratio (the total standard&poor 500 company earnings divided by standard&poor price level), New York Stock Exchange (NYSE) share volume/price (the NYSE trading volume in a month divided by NYSE price level), Credit spread BAA–AAA (the difference between yields on BAA and AAA rated corporate bonds), Treasury-bill (the yield of 3-month maturity treasury bills), Change in T-bill (the change in the yield of 3-month maturity treasury bills from the previous period), Term spread (the difference between the yield on 10-year maturity treasury bond and 3-month maturity treasury bills), yield spread (the difference between the Fed Fund Rate and the 3-month treasury bill), inflation (the producer price index), change in inflation (the change in producer price index from the last period), CPI change (the change in consumer price index), and IP 12m log (the logarithm of change in industrial production).

The dataset used in our study ranges from March, 1954 to March, 2005, and contains 613 observations. Based on this dataset, we consider three time horizons, 3-, 5-, and 8-year periods, so that the resulting data are not too short to fit the linear regression model. It is worthwhile noting that we do not adopt the long time horizon (e.g., 10- or 20-year period) primarily because both academics and practitioners suspect that stock returns are unstable and long-term forecasts are meaningless.

3 Results and discussion

We evaluate the performance of candidate models in two aspects: the out-of-sample forecast of the hit rate and the portfolio performance based on the forecast. In the first task, we obtain the forecast of the next-month return and then compare the sign of the predicted return with the sign of the true return in the following month. A model with a relatively higher hit rate demonstrates a stronger ability to forecast the future moves of U.S. stock market. To achieve the second task, we imitate the decision-making process that portfolio managers typically perform when choosing an investment from two asset classes (i.e., stocks vs. the safer investment alternative of U.S. Treasury bonds).

Table 1 Forecast accuracy (hit rate)

	BIC (I)	Averaging (II)	Mixing (III)	Averaging vs. BIC (II) − (I)	t stat	Mixing vs. BIC (III) − (I)	t stat
3-year	0.570	0.620	0.626	0.050	1.742	0.056	1.924
5-year	0.548	0.626	0.624	0.078	2.691	0.076	2.627
8-year	0.530	0.605	0.611	0.075	2.594	0.081	2.797

Table 2 Portfolio performance (monthly difference)

	BIC (I)	Averaging (II)	Mixing (III)	Averaging versus BIC (II) − (I)	t stat	Mixing versus BIC (III) − (I)	t stat
3-year	0.010	0.009	0.008	−0.001	−1.057	−0.002	−1.252
5-year	0.007	0.009	0.009	0.002	1.587	0.002	0.940
8-year	0.008	0.007	0.007	−0.001	−1.038	−0.001	−0.470

Furthermore, we assume that a portfolio manager's choice of investing in the stock market or bond market is entirely dependent on the model prediction for stock market returns in the next month. By comparing the forecasted stock market returns with zero, the portfolio manager decides how to invest in the next month. Specifically, if the forecasted return of the stock market is greater than zero, then the manager invests totally in the stock market (represented by the CRSP value-weighted index returns). Otherwise, the manager invests entirely in U.S. treasury bonds (represented by the returns of the 30-year U.S. treasury bond).

Table 1 shows that all three methods accurately forecast the direction of next-month returns in more than half of the occasions. For example, the BIC method based on 8-year estimation period yields accurate directional forecasts for 53% of the months, while the mixing approach based on a 3-year estimation period leads to accurate directional forecasts for 62.6% of the months. More interestingly, the averaging and mixing approaches offer better predictions than BIC. For the 3-year time period, the averaging and mixing methods, respectively, produce correct forecasts for 62 and 62.6% of the hit rates, and they are 5.0 and 5.6% higher than the BIC approach. Table 1 also demonstrates that the improvements of averaging and mixing are significant at the 10% significance level in the 3-year period and at the 5% significance level in the 5- and 8-year periods. Moreover, the improvements are greater for the 5- and 8-year horizons: both approaches yield about 7.8% better forecast than that of BIC. In sum, the averaging and mixing methods are superior to BIC in out-of-sample predictions of hit rates.

In the context of portfolio management and financial markets, it is noteworthy that one cannot conclude the preferred selection method solely based on the out-of-sample hit rate. In reality, the essential point is whether one method can generate greater investment returns than its competitors. To this end, we form portfolios via the forecasts obtained from the three selection methods. In particular, we assume that the portfolio manager strictly follows the trading rule to manage the portfolio. First, the manager uses the historical information to linearly regress the future stock market return on the predictive variables across three time horizons (3-, 5-, and 8-year periods). Based on the resulting parameter estimates, he/she next generates the forecast return for the coming month. Finally, the manager invests his/her portfolio entirely in the stock market in the coming month if the forecast return is greater than zero, and entirely in U.S. treasury bonds if the forecast return is smaller than (or equal to) zero.

Table 2 shows that, despite the superior direction prediction obtained from the averaging and mixing approaches, portfolios based on such methods do not lead to a statistically better performance (and sometimes produce insignificantly worse performance) than the portfolio based on BIC. Consequently, our analysis reveal that there is no single model selection method capable of outperforming other methods in portfolio performance.

4 Conclusion

In this study, we evaluate three model selection techniques within the context of financial markets. Although the more recent techniques, averaging and mixing, demonstrate better ability to predict the direction of future stock market movement, portfolios based on such forecasts do not result in better performance than those obtained from the Bayesian selection criterion, BIC. In addition to model mixing and model averaging, we also employed Tibshirani (1996) least absolute shrinkage and selection operator (Lasso) approach to study the performance of hit rate and portfolio. Because the results are not very encouraging, we do not present them here. In sum, despite the recent proliferation of the sophisticated model-selection techniques, BIC remains an easy and effective model selection criterion for the purpose of forecasting stock market returns.

Acknowledgments We are grateful to the Editor and the referee for their helpful comments. Wang's research is supported in part by a NSFC grant (No. 10771006).

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov, F. Csaki (Eds.), *2nd International symposium on information theory* (pp. 267–281). Budapest: Akademia Kiado.

Akaike, H. (1978a). A Bayesian analysis of the minimum AIC procedure. *Annals of the Institute of Statistical Mathematics, 30*, 9–14.

Akaike, H. (1978b). On the likelihood of a time series model. *The Statistician, 27*, 217–235.

Akaike, H. (1979). A Bayesian extension of the minimum AIC procedure of autoregressive model fitting. *Biometrika, 66*, 237–242.

Ang, A., Bekaert, G. (2007). Stock return predictability: is it there? *Review of Financial Studies, 20*, 651–707.

Avramov, D. (2002). Stock return predictability and model uncertainty. *Journal of Financial Economics, 64*, 423–458.

Black, F. (1993). Estimating expected returns. *Financial Analyst Journal, 49*, 36–38.

Bossaerts, P., Hillion, P. (1999). Implementing statistical criteria to select return forecasting models: what do we Learn? *Review of Financial Studies, 12*, 405–428.

Buckland, S. T., Burnham, K. P., Augustin, N. H. (1997). Model selection: an integral part of inference. *Biometrics, 53*, 603–618.

Burnham, K. P., Anderson, D. R. (2002). *Model selection and multimodel inference* (2nd edn.). New York: Springer.

Campbell, J. (1987). Stock returns and the term structure. *Journal of Financial Economics, 18*, 373–399.

Chen, N. F., Roll, R., Ross, S. (1986). Economic forces and the stock market. *Journal of Business, 59*, 383–403

Cremers, M. (2002). Stock return predictability: a Bayesian model selection perspective. *Review of Financial Studies, 15*, 1223–1249

George, E. I., McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association, 88*, 881–889.

Hastie, T., Tibshirani, R., Friedman, J. (2001). *The elements of statistical learning*. New York: Springer.

Hjort, N. L., Cleaskens, G. (2003). Frequentist model average estimators (with discussion). *Journal of the American Statistical Association, 98*, 879–899.

Hoeting, J., Madigan, D., Raftery, A., Volinsky, C. (1999). Bayesian model averaging. *Statistical Science, 14*, 382–401.

Leamer, E. E. (1978). *Specification searches: ad hoc inference with nonexperimental data*. New York: John Wiley.

Lo, A., MacKinlay, A. (1997). Maximizing the predictability in the stock and bond markets. *Macro Economics Dynamics, 1*, 102–134.

Malkiel, B. G. (2004). Models of stock market predictability. *Journal of Financial Research, 27*, 449–459.

Peart, S. (1996). *The economics of W. S. Jevons (Routledge studies in the history of economics, 9)*. New York: Springer.

Rachev, S. T., Mittnik, S., Fabozzi, F. J., Focardi, S. M,, Jasic, T. (2006). *Financial econometrics: from basics to advanced modeling techniques*. New York: Wiley.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*, 461–464.

Tibshirani, R. J. (1996). Regression shrinkage and selection via lasso. *Journal of the Royal Statistical Society, Series B, 58*, 267–288.

Yang, Y. (2001). Adaptive regression by mixing. *Journal of the American Statistical Association, 96*, 574–588.

Yang, Y. (2003). Regression with multiple candidate models: selecting or mixing? *Statistica Sinica, 13*, 783–809