Combining time series models for forecasting

Hui Zou, Yuhong Yang

Content

- Introduction
- Some preliminaries
- Algorithm AFTER for combining forecasts
- Simulations
- Data Examples
- Conclusion

Introduction

- It’s important to choose the ‘best’ model among a variety of candidates
- However, model selection is often unstable and may cause an unnecessarily high variability in the final estimation/prediction
- Models close to each other are usually hard to distinguish and the model selection criterion values are usually quite close to each other
- With an appropriate weighting scheme, the combined forecast has a smaller variability so that the forecasting accuracy can be improved relative to the use of a selection criterion (when model selection is stable, combining does not necessarily lead to any improvement)

Some preliminaries

- Some model selection criteria
    - AIC,BIC,HQ……
    - form:-log(maximized likelihood)+penalty
- Problems about model selection
    - Identifying the true model is not necessarily optimal for forecasting
    - Model selection can be very unstable

Some preliminaries

- Evaluation of forecasting accuracy

  - average net mean square error in prediction （simulation）

$$ANMSEP(\delta, n_0, n) = \frac{1}{n - n_0 + 1} \sum_{i=n_0+1}^{n+1} E(m_i - \hat{Y}_i)^2$$

where $m_i = E\left(Y_i \middle| Y^{i-1}\right)$, $Y^{i-1} = \{Y_j\}_{j=1}^{i-1}$

  - average square error in prediction(based on real data)

$$ASEP(\delta, n_0, n) = \frac{1}{n - n_0} \sum_{i=n_0+1}^{n} (Y_i - \hat{Y}_n)^2$$

Some preliminaries

- Identifying the true model is not necessarily optimal for forecasting

Example1

Model 1: $Y_n = e_n \to \hat{Y}_{n+1} = 0$

Model 2: $Y_n = \alpha Y_{n-1} + e_n \to \hat{Y}_{n+1} = \hat{\alpha}_n Y_n$

when α is actually nonzero, the ratio of NMSEP

$$\frac{\alpha^2 E{Y_n}^2}{E\ (\hat{\alpha}_n - \alpha)\ ^2{Y_n}^2}$$

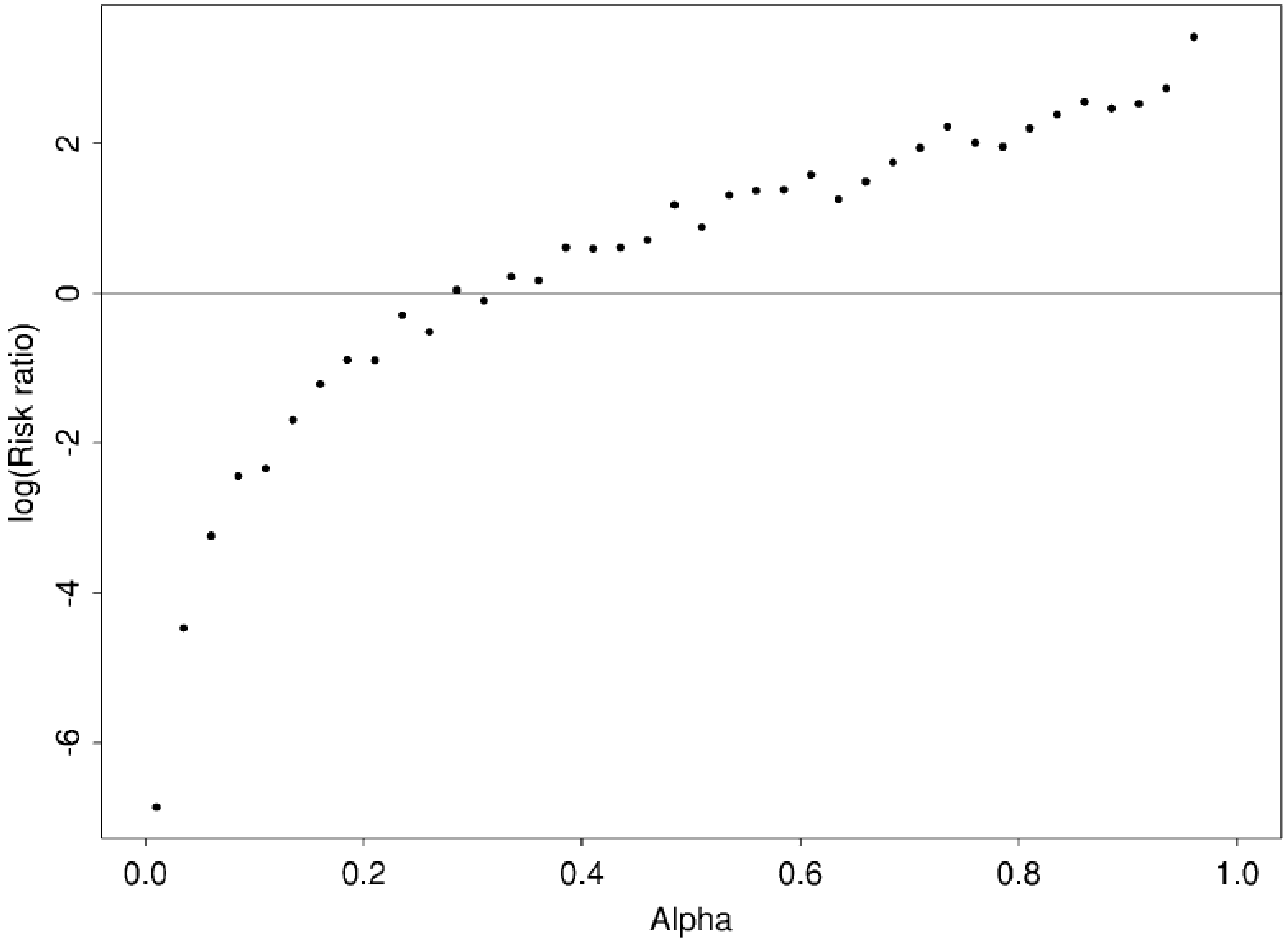

Fig. 1. Comparing true and wrong models in prediction.

Some preliminaries

- Measuring the stability of model selection methods
  - Sequential stability
  - Perturbation stability
    a. Stability in selection
    b. Instability in forecasting

Some preliminaries

- Sequential stability

Let L be an integer between 1 and n-1.; j between n-L and n-1

$$K = \frac{\sum_{j=n-L}^{n-1} I\{\hat{k}_j = \hat{k}_n\}}{L}$$

where $\hat{k}_n$ is the model based on all observations, $\hat{k}_{\mathrm{j}}$denote the model based on $\{Y_i\}_{i=1}^{j}$

Some preliminaries

- Sequential stability

Example 2

Consider AR models. The true model is AR(2) with coefficients(0.278, 0.366). For n=50 and L=20, based on 1000 replications, the average sequential stability of AIC,BIC, HQ and AICc are 0.70, 0.79, 0.75 and 0.72

Some preliminaries

- Perturbation stability

$$\Phi(B)Y_i = \theta(B)e_i \rightarrow \widehat{\Phi}(\mathrm{B}), \hat{\theta}(\mathrm{B})$$

Generate a time series $W_i, \widehat{\Phi}(\mathrm{B})W_i = \hat{\theta}(\mathrm{B})\mu_i$,where $\mu_i \sim N(0, \tau^2\sigma^2)$

Creative new data $\tilde{Y}_i = W_i + Y_i$

Some preliminaries

- Stability in selection
- Instability in forecasting

$$\frac{|\tilde{y}_{i+1} - \hat{y}_{n+1}|}{\hat{\sigma}}$$

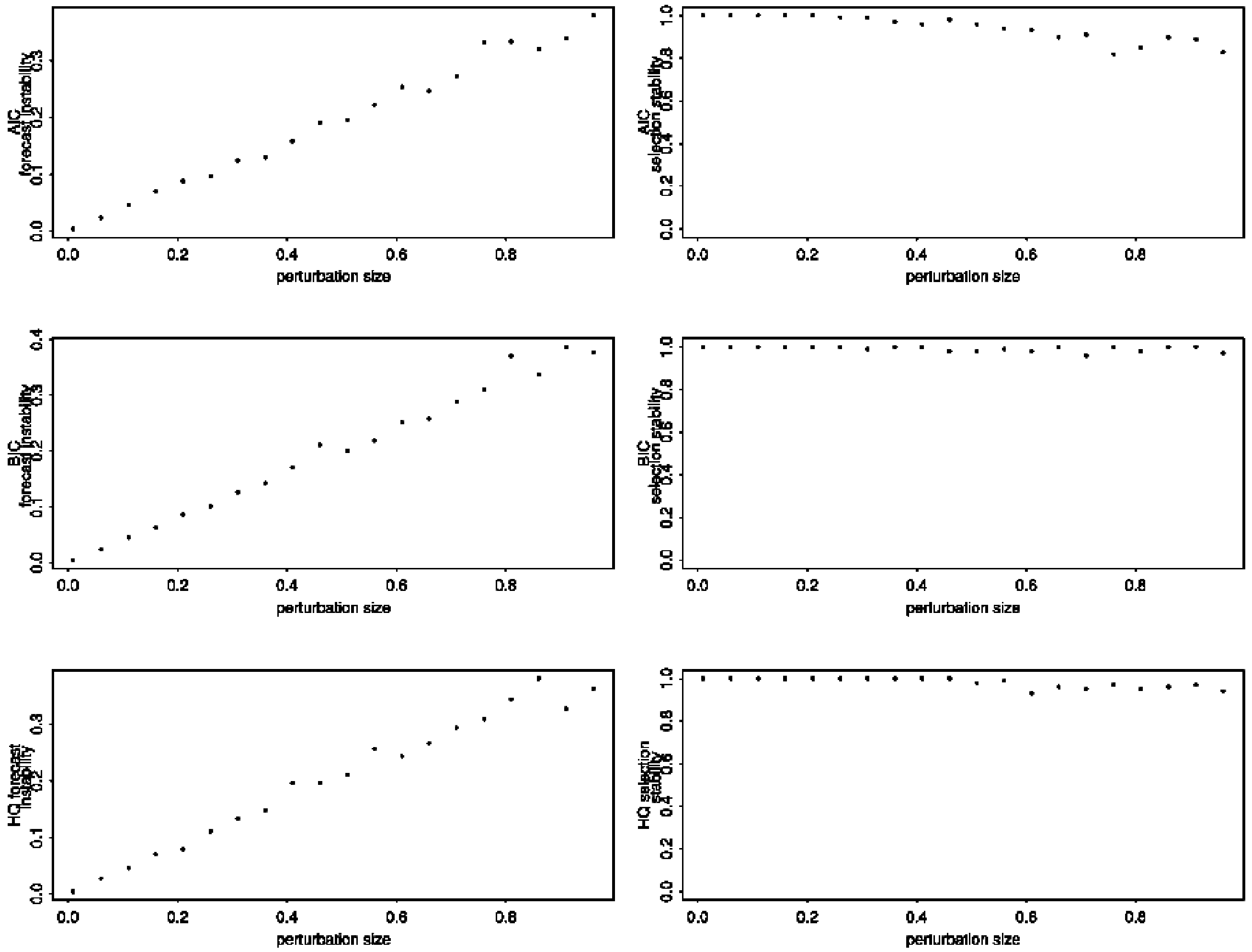

Fig. 2. Forecast and selection stability for data set 1.

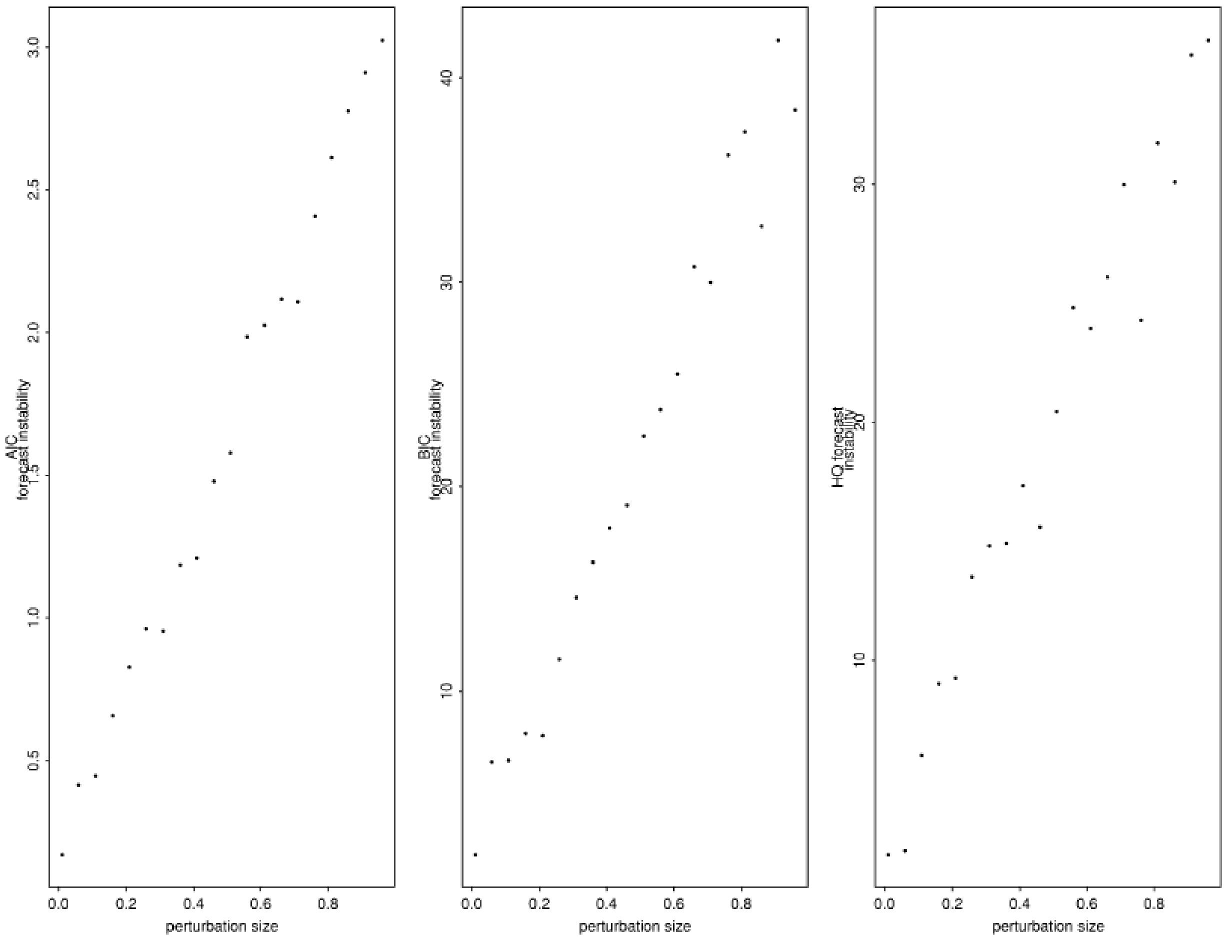

Fig. 3. Forecast instability for data set 3.

Algorithm AFTER for combining forecasts

j indicate forecasting procedure, *i* indicate period

$W_{j,1}$=1/J and for n⩾2

$$W_{j,n}=\frac{\prod_{i=1}^{n-1}\hat{v}_{j,i}^{-1/2}\exp\{-\frac{1}{2}\sum_{i=1}^{n-1}[(Y_i-\hat{y}_{j,i})^2/\hat{v}_{j,i}]\}}{\sum_{j'=1}^{J}\prod_{i=1}^{n-1}\hat{v}_{j',i}^{-1/2}\exp\{-\frac{1}{2}\sum_{i=1}^{n-1}[(Y_i-\hat{y}_{j',i})^2/\hat{v}_{j',i}]\}}. \qquad \hat{y}_n^*=\sum_{j=1}^{J}W_{j,n}\hat{y}_{j,n}$$

$$W_{j,n}=\frac{W_{j,n-1}\hat{v}_{j,n-1}^{-1/2}\exp\{-[(y_{n-1}-\hat{y}_{j,n-1})^2/2\hat{v}_{j,n-1}]\}}{\sum_{j'=1}^{J}W_{j',n-1}\hat{v}_{j',n-1}^{-1/2}\exp\{-[(y_{n-1}-\hat{y}_{j',n-1})^2/2\hat{v}_{j',n-1}]\}}$$

Simulation

- Two simple model（TRUE & WRONG）

The following is a result from a simulation study with n=20 based on 1000 replications. Table 1 gives percentages of selecting the true model by AIC, BIC, HQ and AICc at three representative α values.

Table 1
Percentage of selecting the true model

	AIC	BIC	HQ	AICc
$\alpha = 0.01$	0.157	0.094	0.135	0.144
$\alpha = 0.50$	0.580	0.463	0.557	0.568
$\alpha = 0.91$	0.931	0.884	0.921	0.924

Simulation

Table 2 presents the ANMSEPs of the true model, the wrong model, AIC, BIC, HQ, AICc, and AFTER

Table 2
Model selection vs. combining for the two model case in $ANMSEP(\delta, 12, 20)$ and $NMSEP(\delta, 20)$

	True	Wrong	AIC	BIC	HQ	AICc	AFTER
$\alpha = 0.01$							
$n_0 = 12$	0.116	0.000	0.036	0.025	0.036	0.033	0.029
	(0.03)	(0.000)	(0.003)	(0.002)	(0.003)	(0.002)	(0.001)
$n_0 = 20$	0.094	0.000	0.028	0.017	0.025	0.028	0.027
	(0.005)	(0.000)	(0.004)	(0.003)	(0.004)	(0.004)	(0.003)
$\alpha = 0.50$							
$n_0 = 12$	0.158	0.339	0.211	0.234	0.211	0.216	0.158
	(0.005)	(0.007)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)
$n_0 = 20$	0.124	0.332	0.168	0.192	0.173	0.170	0.156
	(0.008)	(0.015)	(0.011)	(0.012)	(0.011)	(0.011)	(0.010)
$\alpha = 0.91$							
$n_0 = 12$	0.387	5.057	0.944	1.140	0.934	0.988	0.552
	(0.011)	(0.195)	(0.053)	(0.068)	(0.053)	(0.003)	(0.022)
$n_0 = 20$	0.320	5.019	0.614	0.721	0.615	0.615	0.596
	(0.020)	(0.230)	(0.045)	(0.071)	(0.060)	(0.049)	(0.048)

Simulation

Figs. 4 and 5 compare the model selection criteria with AFTER in terms of NMSEP(δ， 20) and ANMSEP（δ， 12,20）, respectively, based on 1000 replications.

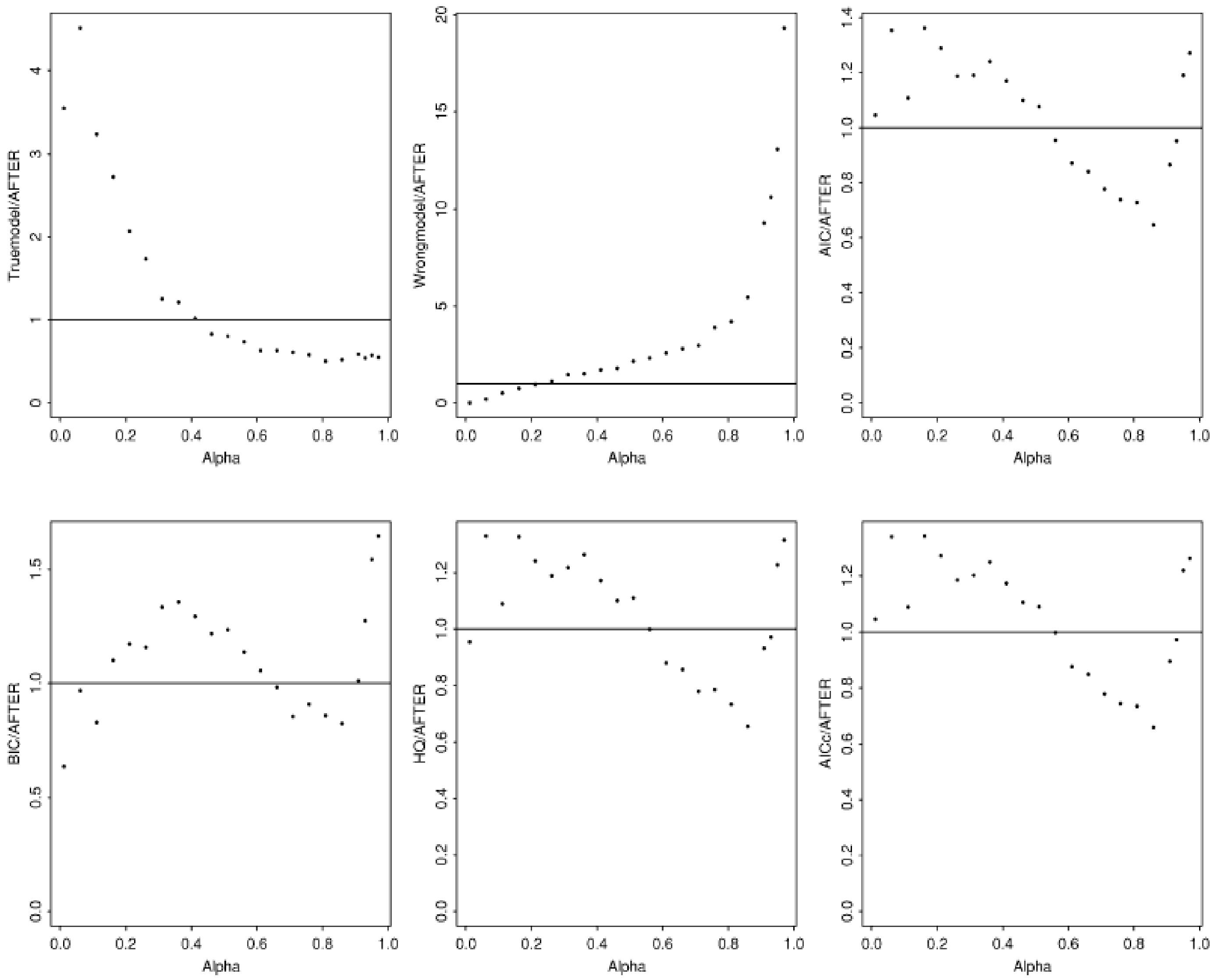

Fig. 4. Comparing model selection with AFTER for the two model case in $NMSEP(\delta, 20)$.

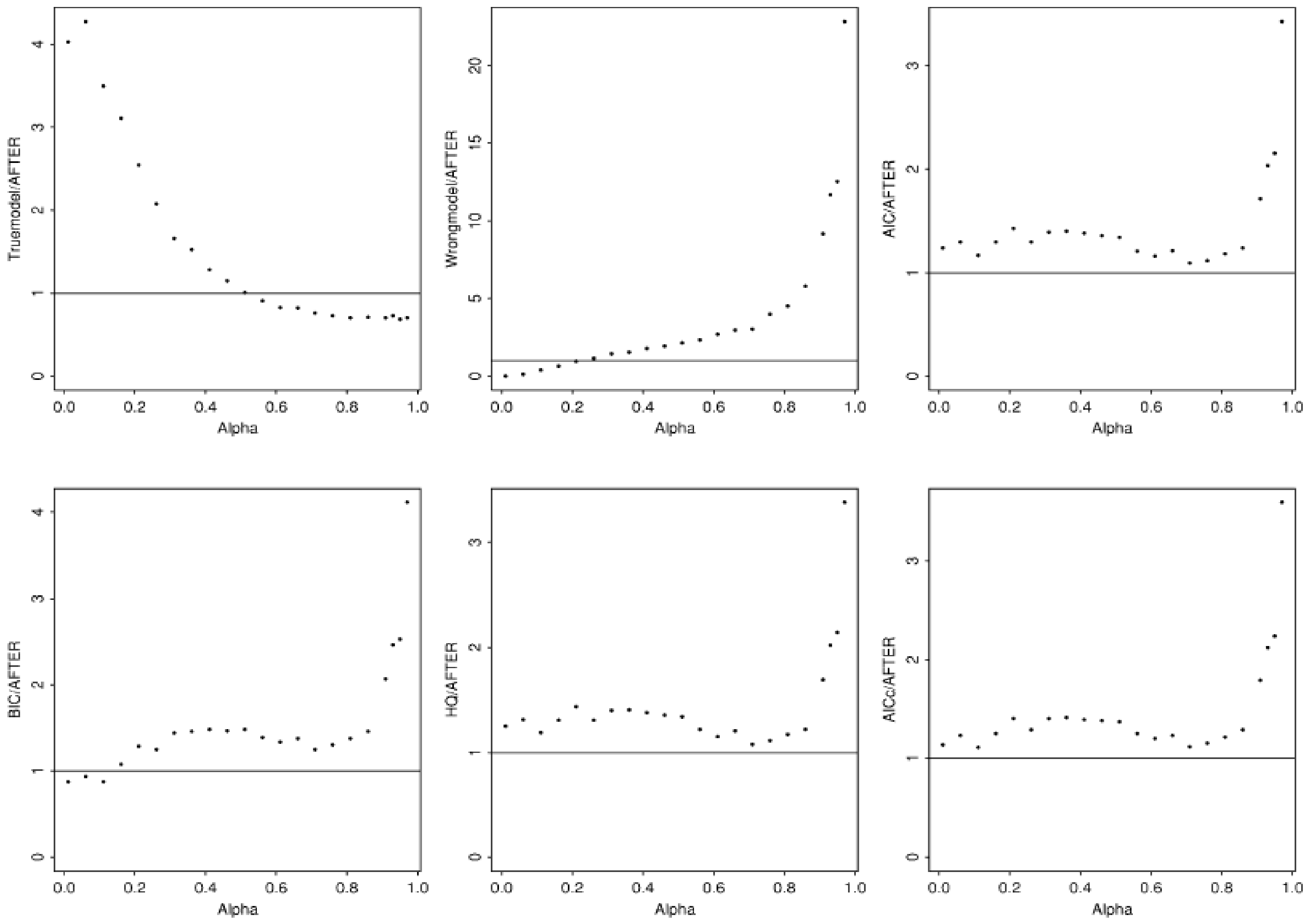

Fig. 5. Comparing model selection with AFTER for the two model case in $ANMSEP(\delta, 12, 20)$.

Simulation

- Table 3 presents the means of the selection instability for α = 0.5 and 0.91 (top and bottom correspond to τ=0.2 and 0.6)

Table 3
Selection stability at two perturbation sizes for three sample sizes

	AIC	BIC	HQ	AICc
$\alpha = 0.5$				
$n = 12$	0.867	0.876	0.863	0.871
	0.778	0.824	0.771	0.794
$n = 20$	0.872	0.856	0.865	0.871
	0.802	0.832	0.805	0.809
$n = 50$	0.959	0.920	0.948	0.962
	0.928	0.883	0.912	0.926
$\alpha = 0.91$				
$n = 12$	0.880	0.874	0.874	0.878
	0.775	0.782	0.780	0.790
$n = 20$	0.957	0.926	0.955	0.954
	0.921	0.888	0.918	0.912
$n = 50$	1.000	1.000	1.000	1.000
	1.000	0.999	1.000	1.000

Simulation

Table 4
Average weights on the models by AFTER

	True	Wrong
$\alpha = 0.01$	0.374	0.626
$\alpha = 0.50$	0.519	0.481
$\alpha = 0.91$	0.793	0.207

Simulation

- AR models with different orders（5 case,8candidate）

  Table 5 shows ANMSEP of selection model and combining model

Table 5
Comparing model selection to combining with AR models

	AIC	BIC	HQ	AFTER	AFTER2	SA
Case 1						
$n_0 = 20$	0.164	0.139	0.151	0.161	0.145	0.210
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
$n_0 = 50$	0.107	0.084	0.089	0.092	0.084	0.131
	(0.007)	(0.006)	(0.006)	(0.006)	(0.005)	(0.007)
Case 2						
$n_0 = 20$	0.162	0.174	0.163	0.139	0.139	0.149
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
$n_0 = 50$	0.103	0.122	0.105	0.097	0.097	0.104
	(0.005)	(0.007)	(0.006)	(0.005)	(0.005)	(0.005)
Case 3						
$n_0 = 20$	0.167	0.163	0.164	0.137	0.137	0.149
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
$n_0 = 50$	0.121	0.131	0.125	0.097	0.102	0.098
	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)
Case 4						
$n_0 = 20$	0.194	0.223	0.205	0.152	0.166	0.137
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)
$n_0 = 50$	0.141	0.184	0.157	0.128	0.141	0.102
	(0.007)	(0.008)	(0.008)	(0.006)	(0.006)	(0.005)
Case 5						
$n_0 = 20$	0.813	0.865	0.835	0.745	0.767	0.833
	(0.012)	(0.013)	(0.013)	(0.012)	(0.012)	(0.014)
$n_0 = 50$	0.604	0.771	0.681	0.593	0.658	0.606
	(0.039)	(0.042)	(0.039)	(0.034)	(0.036)	(0.034)

Simulation

- Random models

select the true AR order (uniformly between 1 and 8) and then randomly generate the coefficients with uniform distribution on [-10,10].110 true model

Simulation

Table 6 compares AFTER with AIC, BIC, and HQ by examining the ratio of the net risks for n = 100 and n_0 = 50 based on 100 replications for each model

Table 6
Comparing model selection to combining with AR models: random case

	AIC	BIC	HQ
Median loss ratio	1.347	1.286	1.247
Risk ratio	1.663	1.678	1.578
	(0.092)	(0.133)	(0.099)

Simulation

Fig.6 and Fig.7 show box plots of the risks of AIC, BIC, HQ and AFTER,and the risks of AIC, BIC, and HQ relative to that of AFTER from the 110 random models

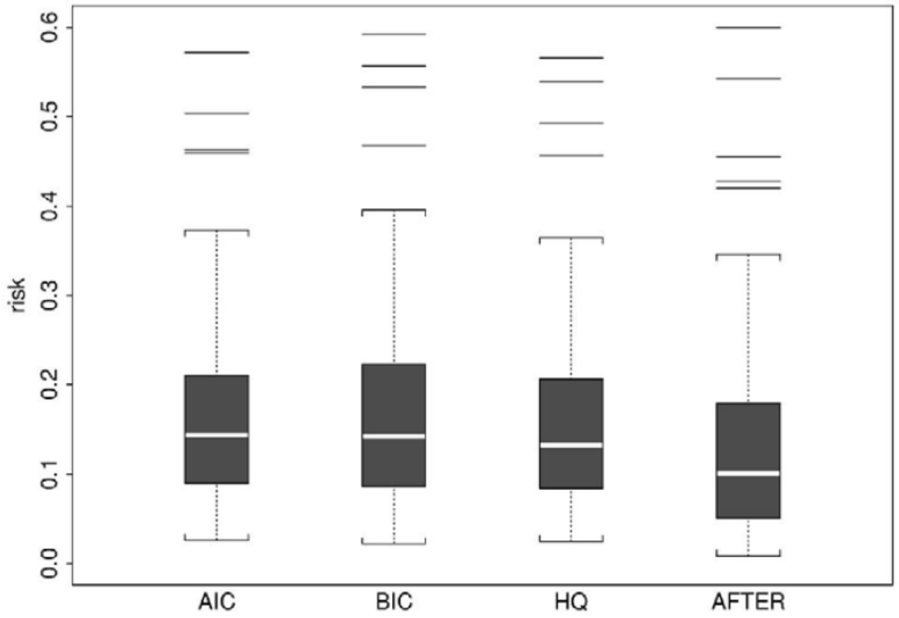

Fig. 6. Risks of AIC, BIC, HQ and AFTER with random AR models.

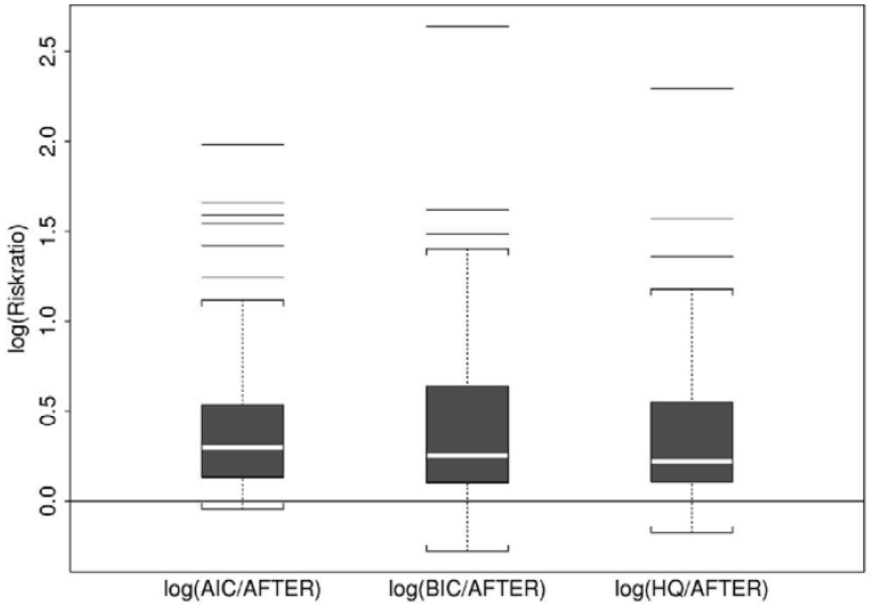

Fig. 7. Risk ratios of AIC, BIC, and HQ compared to AFTER with random AR models.

Simulation

- ARIMA models with different orders

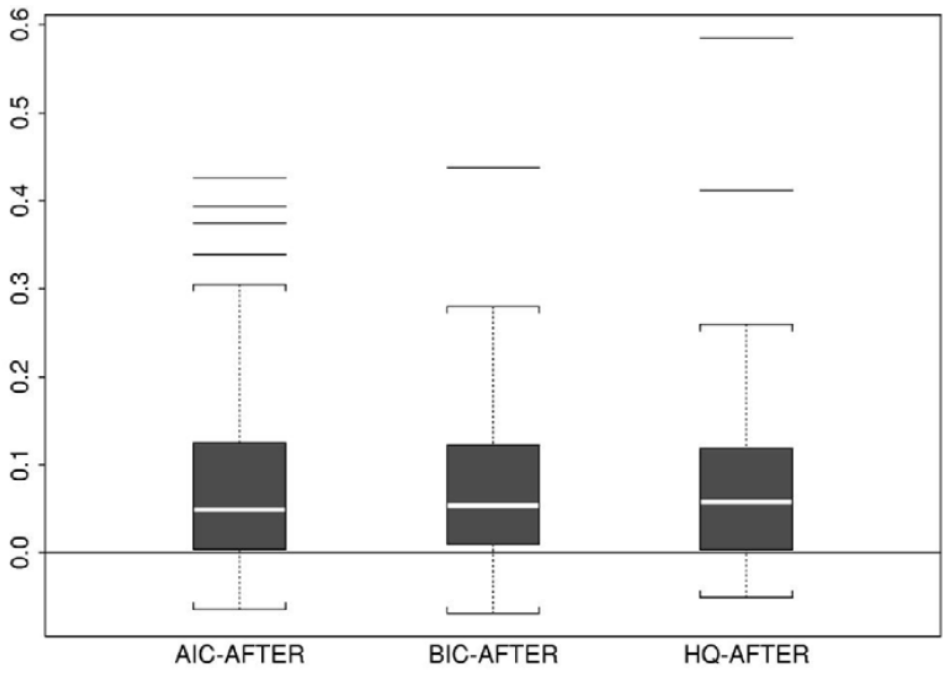

Fig. 8. Loss differences of AIC, BIC, and HQ compared to AFTER for an ARIMA model.

Data examples

- Data set 1: daily average number of defects per truck at the end of the assembly line in a manufacturing plant （n=45）
- The ASEP(n_0, n) for AIC, BIC, and HQ are all equal to 0.249 and is 0.254 for AFTER. The sequential stability K(L = 15) is 1 for all the selection methods and they all choose AR(1).

Data examples

- Data set 2:levels of Lake Huron in feet for July from 1875 to 1972 (n=98)

True model ARMA(1,1),candidate models ARIMA(p,d,q), p = 0,1,2, d = 0,1, q = 0,1,2

- For n_0 = 78, the ASEP(n_0, n) for AIC, BIC, and HQ is 0.721, 0.665,and 0.665, respectively. For AFTER, the ASEP(n_0, n) is 0.628, a reduction of 13, 6 and 6%, respectively, compared to the model selection criteria.

Data examples

- Data set 3: Australian clay brick monthly production statistics (n=155)

ARIMA(p,d,q) models with p,q = 0,…,5, d = 0,1 are considered as candidate models here

- The sequential instability K (with L = 60) for AIC, BIC and HQ is 0.41, 0.26, and 0.17, respectively, suggesting that there is substantial selection instability in the data. The ASEPs for AIC, BIC, HQ, and AFTERare 704.2, 813.4, 785.2, and 635.3, respectively. Note that AIC, BIC and HQ have 11, 28 and 23% higher error compared to AFTER, respectively.

Conclusion

- Model selection can outperform AFTER when there is little difficulty in finding the best model by the model selection criteria.
- When there is significant uncertainty in model selection, AFTER tends to perform better or much better in forecasting than the information criteria AIC, BIC and HQ.
- The proposed instability measures seem to be sensible indicators of uncertainty in model selection and thus can provide information useful for assessing forecasts based on model selection

Combining forecasts: Performance and coherence

Main idea:

The relative error reduction in a combined forecast depends upon the extent to which the component forecasts contain unique/independent information