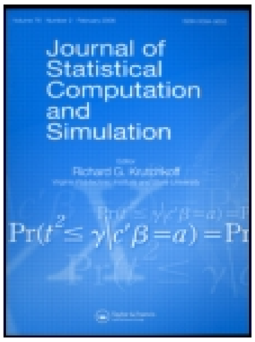

Journal of Statistical Computation and Simulation

ISSN: 0094-9655 (Print) 1563-5163 (Online) Journal homepage: https://www.tandfonline.com/loi/gscs20

Empirical information criteria for time series forecasting model selection

Baki Billah , Rob J. Hyndman & Anne B. Koehler

To cite this article: Baki Billah , Rob J. Hyndman & Anne B. Koehler (2005) Empirical information criteria for time series forecasting model selection, Journal of Statistical Computation and Simulation, 75:10, 831-840, DOI: 10.1080/00949650410001687208

To link to this article: https://doi.org/10.1080/00949650410001687208

Published online: 16 Aug 2006.

Journal of Statistical Computation and Simulation
Vol. 75, No. 10, October 2005, 831–840

Empirical information criteria for time series forecasting model selection

BAKI BILLAH†, ROB J. HYNDMAN*† and ANNE B. KOEHLER‡

†Department of Econometrics and Business Statistics, Monash University, Clayton, Vic 3800, Australia
‡Department of Decision Sciences and Management Information Systems, Miami University, Oxford, OH 45056, USA

(Revised 11 June 2003; in final form 10 February 2004)

In this article, we propose a new empirical information criterion (EIC) for model selection which penalizes the likelihood of the data by a non-linear function of the number of parameters in the model. It is designed to be used where there are a large number of time series to be forecast. However, a bootstrap version of the EIC can be used where there is a single time series to be forecast. The EIC provides a data-driven model selection tool that can be tuned to the particular forecasting task.

We compare the EIC with other model selection criteria including Akaike's information criterion (AIC) and Schwarz's Bayesian information criterion (BIC). The comparisons show that for the M3 forecasting competition data, the EIC outperforms both the AIC and BIC, particularly for longer forecast horizons. We also compare the criteria on simulated data and find that the EIC does better than existing criteria in that case also.

Keywords: Exponential smoothing; Forecasting; Information criteria; M3 competition; Model selection

JEL Classification: C53; C52; C22

1. Introduction

In many industrial applications, a large number of series need to be forecast on a routine basis. The forecaster may either select one appropriate model for all series under consideration, or may use a general selection methodology which will select the appropriate model for each series from a group of competitive models. The appropriate choice of forecasting model has the potential for major cost savings through improved accuracy.

Suppose we have a single time series of length n and N possible models from which to choose. We can choose amongst these models using an information criterion (IC), defined as a penalized log-likelihood:

$$\mathrm{IC} = -2\log L(\hat{\theta}) + 2f(n, q), \tag{1}$$

*Corresponding author. Email: rob.hyndman@buseco.monash.edu.au

Journal of Statistical Computation and Simulation
ISSN 0094-9655 print/ISSN 1563-5163 online © 2005 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/00949650410001687208

where $\log L(\hat{\theta})$ is the maximized log-likelihood function, θ is the q-vector of unknown free parameters, and $f(n, q)$ is the corresponding penalty function. The model with the smallest value of IC is the chosen model.

Six commonly-used information criteria are Akaike's information criterion (AIC; [1]), the Bayesian information criterion (BIC; [2]), Hannan and Quinn's criterion (HQ; [3]), Mallows' criterion (MCp; [4]), the generalized cross validation criterion (GCV; [5]), and the finite prediction error criterion (FPE; [6]). The penalty functions of these criteria are as follows:

Criterion	Penalty function $f(n, q)$
AIC	q
BIC	$q\log(n)/2$
HQ	$q\log(log(n))$
MCp	$n\log(1+2q/r)/2$
GCV	$-n\log(1-q/n)$
FPE	$(n\log(n+q)-n\log(n-q))/2$

where $r = n - q^*$ and q^* is the number of free parameters in the smallest model that nests all models under consideration.

Any of these six information criteria may be used for automatic selection among competing forecasting models (*e.g.*, [7]). However, it is not clear, which IC is best for a given forecasting task, and any one IC does not perform well for all forecasting model selection problems (see, for example, [8–10]). We overcome this problem by proposing a data-driven IC that we call the empirical information criterion (EIC). Rather than using a fixed penalty function $f(n, q)$, we estimate the penalty function for the particular forecasting task, using an ensemble of similar time series.

The plan of this article is as follows. Section 2 introduces the EIC. Section 3 describes the application of the EIC to the M3 forecasting competition data using exponential smoothing models, and we show that it performs better than the existing information criteria. Section 4 applies the bootstrap version of the EIC which is applicable when there is only one series to be forecast. The article ends with some concluding remarks in section 5.

2. Empirical information criteria

Suppose we have m time series that are 'similar' to each other. Let $y_{t,j}$ be the tth observation of the jth series ($j = 1, \ldots, m$ and $t = 1, \ldots, n_j$). We denote the jth series by $\boldsymbol{y}_j$, and the ensemble of series by $\boldsymbol{y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m]$.

This situation can arise when we have a large inventory of m products for which sales need to be forecast on a regular basis. Alternatively, we can fit an initial model to the series of interest (chosen using the AIC for example), and then generate m bootstrap series from the fitted model.

We assume that the penalty does not depend on the length of the series. This is not as restrictive as it first appears because we are estimating the penalty function based on time series that are similar. Thus, all m series will usually be of the same or similar length.

The EIC has $f(n, q) = k_q q$, where k_q is the penalty weight for a model with q parameters. Thus

$$\text{EIC} = -2\log L(\hat{\theta}) + 2k_q q.$$

The model with the smallest EIC is the chosen model. If q_i is the number of parameters for the ith model ($i = 1, 2, \ldots, N$), then the penalty weights k_{q_i}, need to be estimated from the ensemble of m series. Without loss of generality, let the first model have the fewest parameters and assume $k_{q_1} = 0$.

Each series in the ensemble is divided into two segments: the first segment consists of $n_j^* = n_j - H$ observations; the second segment consists of the last H observations. The value of H needs to be chosen by the forecaster according to what is appropriate for the particular series of interest. A common choice will be to set H to the largest forecast horizon required. Another common choice will be to set $H = 1$ as multi-step-ahead forecasts can be viewed as iterating one-step-ahead forecasts.

2.1 *Penalty estimation*

We need to select a k_q value for each unique q in $\{q_2, \ldots, q_N\}$ using the m series in $\boldsymbol{y}$. Small changes in k_q will not usually result in a change in the selected model. Therefore, this is not a smooth optimization problem. There is no reason for the values of k_q to remain positive. Consequently, we consider values of k_q between $-2\log(n)$ and $2\log(n)$ in steps of size δ. This range of values is wide enough to contain all of the commonly used penalty functions. We have found that $\delta = 0.25$ works well in practice. Assuming all values of $\{q_1, \ldots, q_N\}$ are unique, and that there are ζ values of k_{q_i} in the grid for each i, then there are ζ^{N-1} possible sets of $\{k_{q_2}, \ldots, k_{q_N}\}$.

The steps for estimating k_q are as follows:

Step 1: Model estimation

(1a) For each of the m series, use the first n_j^* observations to estimate the parameters in each of the N competing models using maximum likelihood estimation.
(1b) Record the maximized log-likelihoods for all estimated models.

Step 2: Penalty estimation

(2a) For each trial set of $k_{q_2}, \ldots, k_{q_N}$, select a model for each time series by using EIC.
(2b) For each set of $k_{q_2}, \ldots, k_{q_N}$ and each forecast horizon h, calculate the mean absolute percentage error (MAPE) across the m time series to obtain

$$\mathrm{MAPE}(h; k_{q_2}, \ldots, k_{q_N}) = \frac{100}{m} \sum_{j=1}^{m} \frac{|y_{n_j^*+h} - \hat{y}_{n_j^*}(h)|}{y_{n_j^*+h}},$$

where $\hat{y}_{n_j^*}(h)$ is the h-step ($h = 1, \ldots, H$) ahead forecast for the model selected for the jth series.
(2c) Select a value of $\{k_{q_2}^{(h)}, \ldots, k_{q_N}^{(h)}\}$ by minimizing MAPE $(h; k_{q_2}, \ldots, k_{q_N})$ over the grid of $k_{q_2}, \ldots, k_{q_N}$. Thus, a set $\{k_{q_2}^{(h)}, \ldots, k_{q_N}^{(h)}\}$ is selected for each forecast horizon h, $h = 1, \ldots, H$.
(2d) A final value of k_{q_i} is obtained by averaging the H values of $k_{q_i}^{(h)}$ as follows:

$$k_{q_i} = \frac{1}{H} \sum_{h=1}^{H} k_{q_i}^{(h)}. \tag{2}$$

We then use the selected set $\{k_{q_1}, \ldots, k_{q_N}\}$ in equation (2) to find the best model for each series $\boldsymbol{y}_j$ (using all n_j observations) and produce forecasts from this chosen model.

We advocate a grid search in these algorithms because the MAPE function is complicated and relatively ill-behaved. However, it does lead to high computational time which increases sharply with the number of parameters and so can be extremely high for small δ. For a large number of parameters, the simulated annealing algorithm of Goffe *et al.* [11] can be used instead.

Variations on the algorithm can be obtained by replacing the MAPE criterion by some other criteria. For example, mean absolute error (MAE), mean square error (MSE), and root mean square error (RMSE) (but note that these three assume all m series are on the same scale).

2.2 *Bootstrap EIC*

The EIC assumes a suitable ensemble of m series to use in calibrating the penalty function. However, frequently only one series will be available. In this case, a bootstrap approach may be used.

Chen *et al.* [12] proposed a bootstrap approach to estimate a suitable penalty function for selecting the order p of an AR(p) model. Grunwald and Hyndman [13] used a similar idea for determining the degree of smoothness in non-parametric regression. We follow a similar approach by generating the additional m series in the ensemble using a bootstrap. We assume that the series to be forecast is stationary. If this is not so, it should be made stationary through transformations and differences.

Let y_0 denote the one series of interest, and let it be of length n. Then the steps for the bootstrap EIC are as follows.

Step 0: Bootstrap sample generation

(0a) Fit a high-order AR(p) model to $\boldsymbol{y}_0$, the series of interest, and calculate the residuals $\boldsymbol{z} = \{z_1, z_2, \ldots, z_n\}$.

(0b) Generate m bootstrap samples of size n from the residuals $\boldsymbol{z}$. Then, generate m samples of size n from the fitted AR(p) model using the m bootstrap samples of residuals as the errors.

Then the EIC can be applied to obtain the optimal penalty functions. These penalty functions can then be applied to $\boldsymbol{y}_0$ to obtain a new model for the series of interest. The candidate models should all be stationary in this case; they need not be restricted to AR models. Because the m series are all on the same scale, we use RMSE instead of MAPE in Step 2 of the algorithm as it is more numerically stable.

3. Example 1: non-seasonal exponential smoothing models and the M3 data

Exponential smoothing methods have been shown to perform very well for forecasting [14–16]. Hyndman *et al.* [7] describe 24 such exponential smoothing methods and provide state space models for each of them. This allows the likelihood of each model to be computed and allows penalized-likelihood model selection to be used. In this application, we apply four linear exponential smoothing models to the 3003 time series that were part of the M3 competition [17]. These are all real economic or business time series that were used to compare a vast range of forecasting methods. The models are described below; in each case e_t is assumed to be a Gaussian white noise.

Model 1. Local level model (LLM): $y_t = \ell_{t-1} + e_t$, where $\ell_t = \ell_{t-1} + \alpha e_t$ is the local level at time t and α is the exponential smoothing parameter. This underpins the simple exponential smoothing (SES) method.

Model 2. Local level model with drift (LLMD): $y_t = \ell_{t-1} + b + e_t$, where $\ell_t = \ell_{t-1} + b + \alpha e_t$ is the local level at time t, b is the drift and α is the exponential smoothing parameter. This underpins the SES with drift method. Hyndman and Billah [18] show that the LLMD is identical to the theta method of Assimakopoulos and Nikolopoulos [19] which performed well in the M3 competition of Makridakis and Hibon [17]. Hence, this method is of considerable interest to forecast practitioners.

Model 3. Local trend model (LTM): $y_t = \ell_{t-1} + b_{t-1} + e_t$, where $\ell_t = \ell_{t-1} + b_{t-1} + \alpha e_t$, $b_t = b_{t-1} + \beta e_t$. Here, b_t is the growth rate with exponential smoothing parameter β. It underpins Holt's method.

Model 4. Damped trend model (DTM): $y_t = \ell_{t-1} + b_{t-1} + e_t$, where $\ell_t = \ell_{t-1} + b_{t-1} + \alpha e_t$, $b_t = \phi b_{t-1} + \beta e_t$, and ϕ is the damped parameter. It underpins damped exponential smoothing. The LTM is a special case of DTM.

The four models have two, three, four, and five parameters, respectively.

The h-step ahead point forecasts for the LLM, LLMD, LTM, and DTM are given, respectively, by

$$\hat{y}_n(h) = \hat{\ell}_n, \tag{3}$$

$$\hat{y}_n(h) = \hat{\ell}_n + h\hat{b}, \tag{4}$$

$$\hat{y}_n(h) = \hat{\ell}_n + h\hat{b}_n, \tag{5}$$

and

$$\hat{y}_n(h) = \hat{\ell}_n + \hat{b}_n \sum_{i=0}^{h-1} \hat{\phi}^i, \tag{6}$$

where $\hat{\ell}_n$, $\hat{b}$, $\hat{b}_n$, and $\hat{\phi}$ are maximum likelihood estimates of ℓ_n, b, b_n, and ϕ, respectively.

3.1 *Calculations and results*

For the annual data in the M3 competition, the above models are used in this article as the competitive models. Previous studies (*e.g.*, [14, 17]) show that for seasonal data, the deseasonalized exponential smoothing methods do better than their corresponding seasonal versions, particularly for monthly data. Therefore, for seasonal data, the deseasonalized versions of these methods are used. The seasonal data are deseasonalized using the ratio-to-moving average method [20], and the forecasts are reseasonalized before calculating the MAPE. (Of course, it is also possible to fit seasonal models and use the EIC to choose between them, but we have not considered that situation here.)

Estimates of parameters are obtained by maximizing the conditional (Gaussian) log-likelihood as described in Ord *et al.* [21]. The likelihood depends on $\boldsymbol{x}_0$ and the parameters α, β, and ϕ. Constrained optimization was employed to obtain the values of $\boldsymbol{x}_0$ and parameters that maximize the log-likelihood conditional on $\boldsymbol{x}_0$.

We treat the annual, quarterly, and monthly data separately. All series in the set of annual time series are used as the ensemble for calibrating the penalty function for these series, and similarly for the quarterly and monthly series.

Each series is divided into two segments: the training set and the test set. The jth time series ($j = 1, \ldots, m$) has n_j observations in the training set and H observations in the test set.

Table 1. Average MAPE for the annual M3 competition data.

	Forecasting horizons						Average	
Methods	1	2	3	4	5	6	1 to 4	1 to 6
AIC	8.2	12.9	22.0	23.8	28.6	29.7	18.8	22.2
BIC	8.1	12.8	21.8	23.3	28.0	29.4	16.5	20.6
HQ	8.2	12.7	21.9	23.6	28.3	29.7	16.6	20.7
MCp	8.2	12.9	22.0	23.8	28.5	29.6	16.7	20.8
GCV	8.1	12.8	21.8	23.5	28.3	29.2	16.6	20.6
FPE	8.2	13.0	22.1	24.0	28.9	30.1	16.8	21.1
EIC	8.4	12.7	21.4	21.4	25.3	25.8	16.0	19.2

Table 2. Average MAPE for the quarterly M3 competition data.

	Forecasting horizons								Average		
Methods	1	2	3	4	5	6	7	8	1–4	1–6	1–8
AIC	5.2	7.9	8.3	9.3	10.5	13.7	13.0	14.1	7.7	9.2	10.3
BIC	5.2	8.0	8.3	9.3	10.5	13.7	13.0	14.0	7.7	9.2	10.3
HQ	5.2	8.0	8.3	9.3	10.5	13.7	13.0	14.2	7.7	9.2	10.3
MCp	5.2	7.9	8.3	9.3	10.5	13.7	13.0	14.2	7.7	9.2	10.3
GCV	5.2	7.9	8.3	9.3	10.5	13.7	13.0	14.1	7.7	9.2	10.3
FPE	5.2	8.1	8.4	9.5	10.9	14.4	13.8	14.9	7.8	9.4	10.6
EIC	5.0	7.8	7.8	8.8	9.5	12.6	11.6	12.6	7.4	8.6	9.5

For annual, quarterly, and monthly data, the values for H are 6, 8, and 18, respectively. The training set is further divided into two subsets. The jth time series has n_j^* observations in the first subset and H observations in the second subset. The data in the training sets are used to estimate the penalty functions (Steps 1 and 2 in section 2) for each series. The penalties are averaged to obtain just one for each of annual, monthly, and quarterly data.

We compare the EIC obtained in this way with the six other criteria outlined in section 1. Each selected model is used to forecast the values in the test set, and the absolute percentage error is computed for each forecasting horizon.

The MAPEs from the M3 competition are presented in tables 1–3. The results show that the EIC performs better than all existing information criteria. Among the existing information criteria BIC is the best. The performances of the criteria AIC, BIC, HQ, MCp, GCV, and FPE are not the same, particularly for yearly data where, as compared to quarterly and monthly data, the series sizes are usually smaller. The strength of the EIC is that it works well for all model selection problems. The estimated penalty weights are presented in table 4.

Table 3. Average MAPE for the monthly M3 competition data.

	Forecasting horizons								Average			
Methods	1	2	3	4	5	8	12	18	1–4	1–8	1–12	1–18
AIC	15.1	13.9	15.7	18.1	14.7	15.6	16.6	21.9	15.7	15.6	16.0	17.6
BIC	15.1	13.8	15.5	17.8	14.6	15.3	16.1	21.8	15.6	15.4	15.7	17.4
HQ	15.2	13.9	15.7	18.0	14.7	15.6	16.6	21.9	15.7	15.6	16.0	17.6
MCp	15.1	13.9	15.7	18.1	14.7	15.6	16.6	21.9	15.7	15.6	16.0	17.6
GCV	15.1	13.9	15.7	18.1	14.7	15.6	16.6	21.9	15.7	15.6	16.0	17.6
FPE	15.2	13.9	15.7	18.0	14.8	15.6	16.6	21.9	15.7	15.6	16.0	17.7
EIC	15.1	13.9	15.4	17.7	14.5	15.1	15.9	21.4	15.5	15.3	15.6	17.2

Table 4. Estimated weights for the M3 competition data.

	Data types		
Estimated weights	Annual	Quarterly	Monthly
k_{q_2}	1.92	1.75	1.03
k_{q_3}	4.41	3.62	3.47
k_{q_4}	1.42	0.69	1.75

Figure 1 shows the penalty functions for AIC and EIC. The estimated penalties for the EIC are highly non-linear. The non-linear form is similar for all three data types with a maximum at $q = 4$. This consistency demonstrates the stability of our procedure. Since the EIC has much higher values of k_{q_i} for the LTM than the other models, it has high penalty and will be less likely to be chosen than other models.

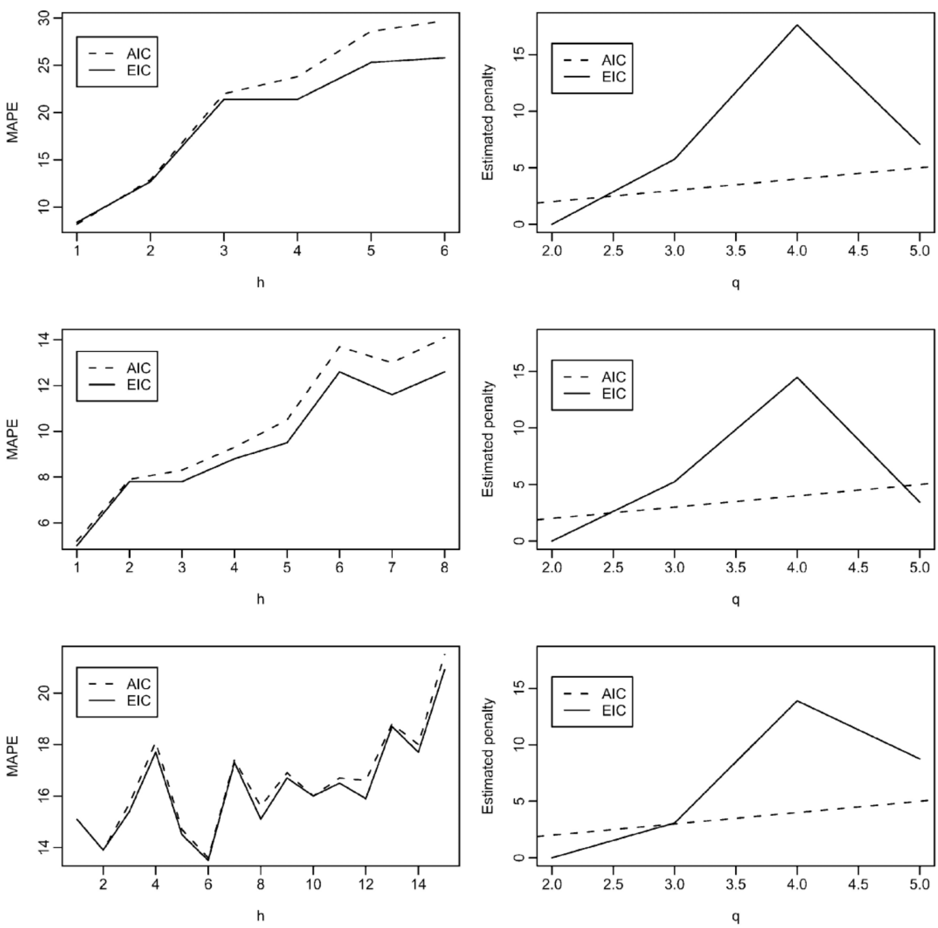

Figure 1. MAPE and estimated penalty functions for EIC.

Table 5. Average RMSE for the simulated data.

		Forecasting horizons						Average	
n	Method	1	2	3	4	5	6	1–4	1–6
20	EIC	1.17	1.95	2.31	2.26	2.24	2.23	1.92	2.03
	AIC	1.13	1.87	2.24	2.35	2.38	2.30	1.90	2.04
30	EIC	1.09	1.72	2.10	2.16	2.09	2.03	1.77	1.87
	AIC	1.18	1.82	2.09	2.17	2.23	2.23	1.82	1.95
50	EIC	1.03	1.65	1.95	2.08	2.15	2.18	1.68	1.84
	AIC	1.10	1.73	2.06	2.26	2.23	2.05	1.79	1.90

Table 6. Estimated average weights for the simulated data.

	Sample size n		
Estimated weights	20	30	50
k_{q_2}	−1.593	−1.514	−1.467
k_{q_3}	−0.833	−0.698	−0.632

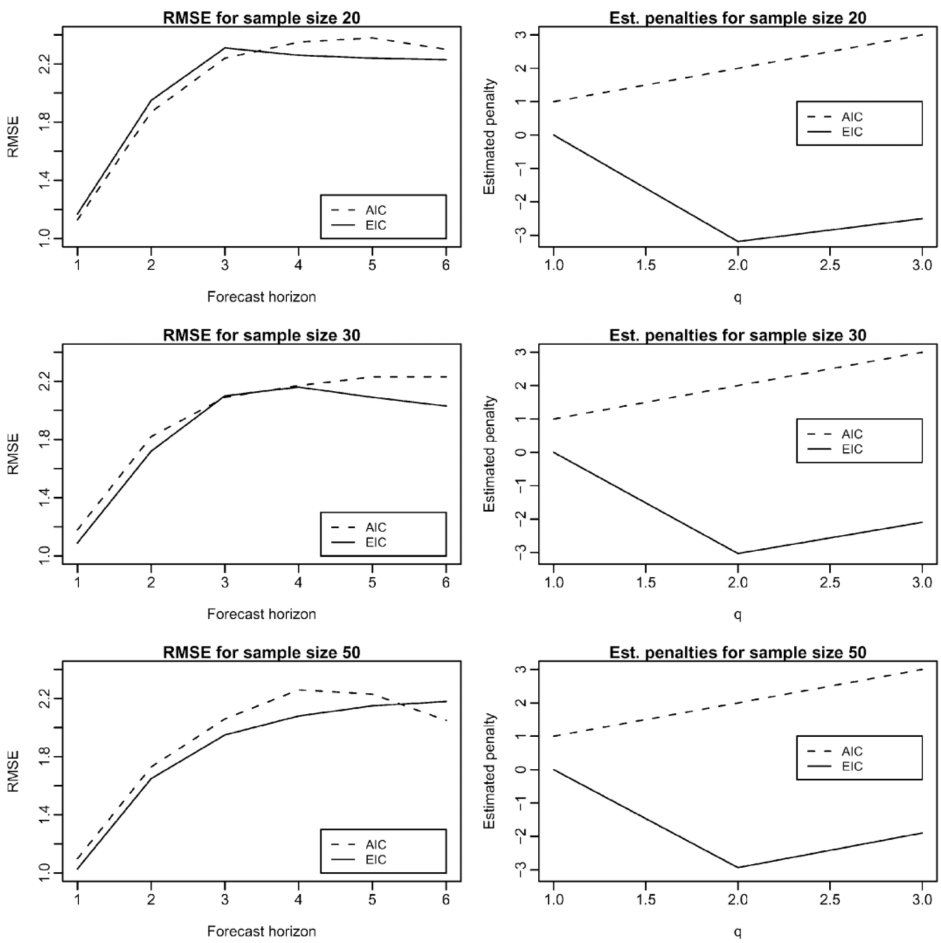

Figure 2. RMSE and estimated penalty functions for AIC and EIC for the simulated data.

4. Example 2: bootstrap EIC applied to simulated data

To test the procedure on simulated data, we generated 500 series from the AR(2) model, $y_t = 1.2y_{t-1} - 0.5y_{t-2} + e_t$, for sample sizes $n = 20$, 30, and 50. The first $n^* = n - 6$ observations were used for estimating the candidate models AR(1), AR(2), and AR(3), and the last $H = 6$ observations were used for computing the RMSE. For each series the penalty values were estimated using the bootstrap EIC. The forecast RMSE for the models selected by AIC and EIC are calculated and presented in table 5. The penalty weights for these simulations are given in table 6 and plotted in figure 2. The results in table 5 show that for larger n, the EIC does substantially better than the AIC. For $n = 20$ there is little difference between the methods.

Again, the non-linear penalty functions (figure 2) are very similar for all sample sizes. This suggests that the penalty functions are determined by the nature of the data (in this case AR(2)), which supports our general philosophy of allowing the entire ensemble of similar data to determine the nature of the penalty function.

5. Conclusions

We have proposed an automatic forecasting model selection algorithm when a large number of series need to be forecast on a routine basis. The methodology is based on a penalized-likelihood criterion and is data-adaptive in the sense that the penalty function is determined by the data to be forecast. Thus, the penalty level is tuned to the attributes of the series being forecast. The EIC has been shown to perform better than all six standard existing information criteria on real and simulated data.

Acknowledgements

This research was supported by the Business and Economics Forecasting Unit, Department of Econometrics and Business Statistics, Monash University, Australia

References

[1] Akaike, H., 1973, Information theory and an extension of the maximum likelihood principle. In: B.N. Petrov and Csaki, F. (Eds) *Second International Symposium on Information Theory* (Budapest: Akademiai Kiado), pp. 267–281.
[2] Schwarz, G., 1978, Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461–464.
[3] Hannan, E.J. and Quinn, B.G., 1979, The determination of the order of an autoregression. *Journal of the Royal Statistical Society, Series B*, **41**, 190–195.
[4] Mallows, C.L., 1964, Choosing variables in a linear regression: a graphical aid. Presented at the Central Regional Meeting of the Institute of Mathematical Statistics, Manhattan, Kansas.
[5] Golub, G.H., Heath, M. and Wahba, G., 1979, Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, **21**, 215–223.
[6] Akaike, H., 1970, Statistical predictor identification. *Annals of Institute of Statistical Mathematics*, **22**, 203–217.
[7] Hyndman, R.J., Koehler, A.B., Snyder, R.D. and Grose, S., 2002, A state space frame-work for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, **18**(3), 439–454.
[8] Hurvich, C.M. and Tsai, C.L., 1991, Bias of the corrected AIC criterion for underfitted regression and time series models. *Biometrika*, **78**, 499–509.
[9] Mills, J.A. and Prasad, K., 1992, A comparison of model selection criteria. *Econometric Reviews*, **11**, 201–233.
[10] Billah, M.B., King, M.L., Snyder, R.D. and Koehler, A.B., 2006, Exponential smoothing model selection for forecasting. *International Journal of Forecasting* (in press).
[11] Goffe, W.L., Ferrier, G.D. and Rogers, J., 1994, Global optimization of statistical functions with simulated annealing. *Journal of Econometrics*, **60**, 65–99.
[12] Chen, C., Davis, R.A., Brockwell, P.J. and Bai, Z.D., 1993, Order determination for autoregressive processes using resampling methods. *Statistics Sinica*, **3**, 481–500.

[13] Grunwald, G.K. and Hyndman, R.J., 1998, Smoothing non-Gaussian time series with autoregressive structure. *Computational Statistics and Data Analysis*, **28**, 171–191.

[14] Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. and Winkler, R., 1982, The accuracy of extrapolation (time series) methods: results of a forecasting competition. *Journal of Forecasting*, **1**, 111–153.

[15] Makridakis, S., Chatfield, C., Hibon, M., Mills, T., Ord, J.K. and Simmons, L.F., 1993, The M2-competition: a real-time judgmentally based forecasting study. *International Journal of Forecasting*, **9**, 5–22.

[16] Fildes, R., Hibon, M., Makridakis, S. and Meade, N., 1998, Generalizing about univariate forecasting methods: further empirical evidence. *International Journal of Forecasting*, **14**, 339–358.

[17] Makridakis, S. and Hibon, M., 2000, The M3-competitions: results, conclusions and implications. *International Journal of Forecasting*, **16**, 451–476.

[18] Hyndman, R.J. and Billah, M.B., 2003, Unmasking the theta method. *International Journal of Forecasting*, **19**, 287–290.

[19] Assimakopoulos, V. and Nikolopoulos, K., 2000, The theta model: a decomposition approach to forecasting. *International Journal of Forecasting*, **16**, 521–530.

[20] Makridakis, S. Wheelwright, S.C. and Hyndman, R.J., 1998, *Forecasting Methods and Applications* (3rd edn) (New York: John Wiley & Sons).

[21] Ord, J.K., Koehler, A.B. and Snyder, R.D., 1997, Estimation and prediction for a class of dynamic nonlinear statistical models. *Journal of the American Statistical Association*, **92**, 1621–1629.