The performance of model selection criteria in the absence of a fixed-dimensional correct model

Erik Olofsen
Department of Anesthesiology, Leiden University Medical Center, The Netherlands
email: e.olofsen@lumc.nl

Objectives

Akaike's information-theoretic criterion (AIC) for model discrimination [1] is often stated to "overfit", *i.e.*, it selects models with a higher dimension than the dimension of the model that generated the data. However, when no fixed-dimensional correct model exists, for example for pharmacokinetic data, AIC (or its bias-corrected version AICc) might be the selection criterion of choice if the objective is to minimize prediction error [2, 3]. The present simulation study was designed to assess the behavior of AIC and other criteria under this type of model misspecification, for various sample sizes and measurement noise levels.

Methods

• M sampling times t_j within $[1/t_{\max}, t_{\max}]$ were chosen according to

$$t_j = \left(\frac{m+1-j}{j}\right)^{\gamma}, \tag{1}$$

with $j = 1 \cdots M$ and $\gamma = \log(t_{\max})/\log(M)$; $t_{\max}$ was set to 100.

• Simulated data were generated by

$$y_j = \frac{1}{t_j}(1+\epsilon_j), \tag{2}$$

where ϵ_j is Gaussian measurement noise with standard deviation σ. A power function of time was chosen because it often fits pharmacokinetic data well, and because it is related to a sum of exponentials [4]:

$$\frac{1}{t} = \int_0^\infty \exp(-\lambda t). \tag{3}$$

A natural approximation of the data y_j is therefore

$$\hat{y}_j(\boldsymbol{\alpha}) = \sum_{i=1}^{N} \alpha_i \exp(-\lambda_i t_j), \tag{4}$$

where M time constants were fixed according to $\lambda_j = 1/t_j$ (without loss of generality), and N of those by were chosen *via* forward selection (see below). The N α_i (denoted by $\boldsymbol{\alpha}$) and σ were estimated using weighted linear least squares [5], with weights $w_j = 1/t_j^2$ in accordance with eq. (2). For any set of M observations y_j (denoted by $\boldsymbol{y}$), there exists a perfect fit only when $N = M$.

• The maximized likelihood associated with $\boldsymbol{y}$ was written as

$$L(\boldsymbol{y}; \hat{\boldsymbol{\alpha}}, \hat{\sigma}) = \prod_{j=1}^{M} \frac{\sqrt{w_j}}{\hat{\sigma}\sqrt{2\pi}} \exp\left(\frac{-w_j(y_j - \hat{y}_j(\hat{\boldsymbol{\alpha}}))^2}{2\hat{\sigma}^2}\right) \tag{5}$$

and the prediction error, associated with a validation set of M observations z_j (denote by $\boldsymbol{z}$), is given by

$$PE(\boldsymbol{z}; \hat{\boldsymbol{\alpha}}) = \frac{1}{M}\sum_{j=1}^{M} w_j(z_j - \hat{y}_j(\hat{\boldsymbol{\alpha}}))^2. \tag{6}$$

• The model selection criteria (MSC) studied were: Akaike's Information-theoretic Criterion (AIC), the corrected version for small sample sizes (AICc), the Bayesian Information Criterion (BIC), the Prediction Error Criterion (PEC) and the likelihood ratio test (LRT) with $P < 0.05$; these are defined by

$$\begin{aligned}
\text{AIC} &\equiv -2\log(L(\hat{\boldsymbol{y}}, \hat{\sigma})) + 2(N+1)\\
\text{AICc} &\equiv -2\log(L(\hat{\boldsymbol{y}}; \hat{\boldsymbol{\alpha}}, \hat{\sigma})) + 2(N+1) + \frac{2(N+1)(N+2)}{M-N-2}\\
\text{BIC} &\equiv -2\log(L(\hat{\boldsymbol{y}}; \hat{\boldsymbol{\alpha}}, \hat{\sigma})) + (N+1)\log(M)\\
\text{PEC} &\equiv PE(\boldsymbol{z}; \hat{\boldsymbol{\alpha}})\\
\text{LRT} &\equiv \text{decrease in} -2\log(L(\hat{\boldsymbol{y}}, \hat{\sigma})) \text{of more than 3.84}
\end{aligned}$$

Methods (continued)

• The steps of the employed forward selection procedure employed are:

1. Select sample size M and noise level σ;
2. Generate training data $\boldsymbol{y}$ and validation data $\boldsymbol{z}$;
3. Set model dimension $N = 1$;
4. Find that λ_j that gives the lowest value of an MSC under study (excluding and fixing the indices of λs found in earlier iterations);
5. Calculate the normalized prediction error PE/σ^2;
6. Go to step 9 if no decrease of the MSC can be found;
7. Set $N = N + 1$;
8. Go to step 4;
9. Go to step 2, repeat 250 times and average model dimension and normalized prediction error.

• The above methods were implemented in the C language; GNU Scientific Library [5] routines were used to generate, store and analyze the data.

Results

• Figures 1 and 2 show the mean normalized prediction errors and model dimensions as functions of the noise level, for sample sizes $M = 15$ and $M = 35$ respectively.

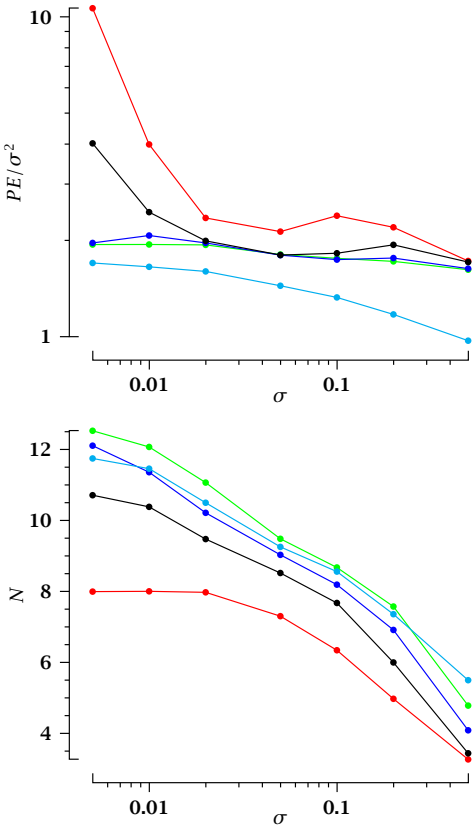

Figure 1: Mean normalized prediction error (PE/σ^2; top panel) and mean model dimension (N; lower panel) for the model selection criteria AIC (green), AICc (red), BIC (blue), LRT (black) and PE (cyan), for sample size $M = 15$.

Results (continued)

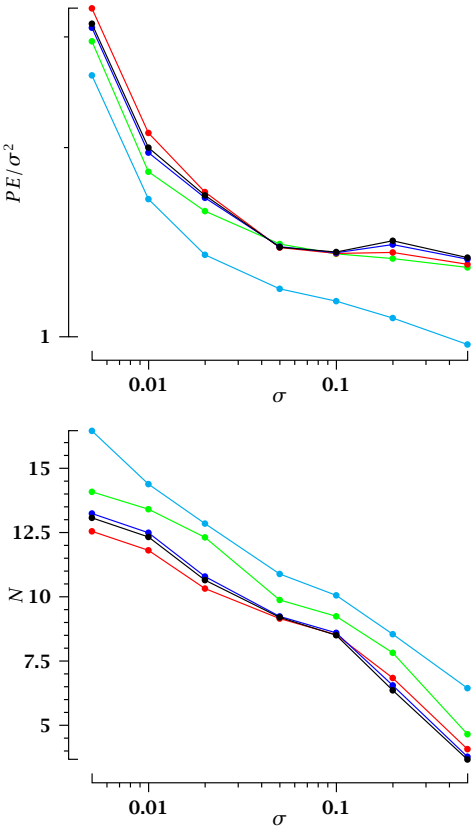

Figure 2: Results for $M = 35$; see legend of figure 1 for explanation.

Conclusions

• From the model selection criteria considered, AIC performed best in the sense that models identified by AIC resulted in the smallest prediction errors.

• AICc did not perform well, especially at low noise levels; this is most likely related to violations of assumptions underlying its derivation.

• When using the prediction error itself as model selection criterion, the forward selection procedure identified models with a smaller prediction error and lower complexity than when using AIC, but only if the number of samples is small. Furthermore, although the models identified with the alternative criteria were smaller, their prediction errors were higher.

• The likelihood ratio test, with $P < 0.05$, resulted in models that were too small to yield optimal prediction performance; using $P < 0.01$ would be even worse.

• Without validation data, the cross-validation prediction error might perform well as a model selection criterion in a forward selection procedure.

References

[1] H Akaike. A new look at the statistical model identification. *IEEE Trans Automat Contr*, 19:716–723, 1974.

[2] J Shao. An asymptotic theory for linear model selection. *Statist Sin*, 7:221–242, 1997.

[3] K P Burnham and D R Anderson. Multimodel inference - understanding AIC and BIC in model selection. *Sociol Meth Res*, 33:261–304, 2004.

[4] K H Norwich. Noncompartmental models of whole-body clearance of tracers: a review. *Ann Biomed Eng*, 25:421–439, 1997.

[5] GSL - GNU Scientific Library. `http://www.gnu.org/software/gsl`, 2007.