Akaike, H. (1987). Factor analysis and AIC (1987). *Psychometrika*, *52*, 317–332 (3,852 citations in Google Scholar as of April, 2016).

In 1986, as the president elect of the Psychometric Society, I was given \$1,000 as a president's discretionary fund to support the Annual Meeting of the Society to be held in Toronto. Without hesitation, I decided to spend the money to bring Professor Akaike from Japan, who was then the director general of the Institute of Statistical Mathematics in Tokyo. He was already famous internationally for his ground breaking idea of using AIC (Akaike Information Criterion) for model selection, and I thought it was opportune to invite him to the meeting as the keynote speaker. Ham Bozdogan (University of Virginia) and I (McGill University) also organized a symposium on the topic, in which several more AIC-related papers were presented. After the meeting, Ivo Molenaar, then the editor of Psychometrika, asked us if we could put together the paper by Akaike and those presented in the symposium as a special section of the journal. The paper by Akaike referred to in the title of this note is the leading paper in the special section published in the September 1987 issue of Psychometrika. It has turned out that this paper is one of the top ten most highly cited papers ever published in the journal. It is interesting to point out that this paper comes as only the third most highly cited paper among all AIC-related papers by Akaike, [1, 2] being the top two, which have attracted four to seven times as many citations as the Psychometrika paper. It is indeed phenomenal that papers dealing with statistical theory have attracted so many citations. The first paper on AIC [1] was subsequently selected for the volume "*Breakthroughs in Statistics*" edited by Kotz and Johnson as one of the most influential papers ever written in statistics. Incidentally, another paper published in the special section [3] has also attracted quite a large number (over 2,000) of citations. This paper emphasized foundational aspects of AIC, while Akaike's paper focussed on applications of AIC to a particular problem in psychometrics.

The high citation rates of these papers are undoubtedly due to the popularity of AIC as a model selection criterion, which in turn is related to its simplicity in use. AIC is defined as

$$\mathrm{AIC}_\pi = -2\log(L_\pi) + 2k, \tag{1}$$

where π indicates a fitted model, L_π is the likelihood of model π maximized over k parameters. It can be readily calculated, once the maximum likelihood of the fitted model is calculated. AIC is evaluated for each model in the set of competing models, and the model associated with the smallest value of AIC is selected as the best-fitting model. The first term of AIC measures the badness-of-fit of the model to the current data set (the data set used to estimate k parameters), which is supplemented by the second term measuring the complexity of the fitted model. In general, the badness-of-fit of the model can be reduced by merely employing more complicated models, which has to be discouraged to increase the predictability of the model for future observations. AIC is deemed to hit a good balance between the two opposing demands. This is achieved by measuring the closeness of the estimated model to the true model. The model chosen as the best fitting model according to the minimum AIC criterion is thus the one closest to the true model among all competing models. See [11] for an illustration of how the exact form of the penalty term, $2k$, emerges in AIC.

In his Psychometrika paper, Akaike proposed to use AIC for determining the number of factors in factor analysis. Factor analysis is considered one of the most important contributions of psychometrics to statistics. It was quite natural that Akaike chose this problem to demonstrate the usefulness of AIC. (Akaike, in fact, admitted that he initially thought of AIC in the context of factor analysis.) At surface, this seemed like a straightforward application of AIC: simply calculate AIC with the number of factors systematically varied, and choose the minimum AIC solution. It turned out, however, that the problem was not as easy as first expected. As he increased the number of factors, he soon encountered improper solutions, in which negative estimates of unique variances were obtained. To overcome the difficulty, Akaike proposed a maximum penalized likelihood (MPL) estimation, in which the (log) likelihood to be maximized was appended by a penalty term to keep the estimates of variances positive. This strategy was successful to some extent. The penalty term, however, had a parameter that regulated the amount of driving force that kept the variance estimates positive. This parameter was not treated as part of the optimization framework. Akaike only tried a few values for the penalty parameter and informally observed the consequences.

This left some arbitrariness in his proposal.

There was also another problem in employing the MPL estimation. The penalty introduced in the MPL estimation affects the definition of AIC itself. It is generally believed that the effective number of parameters in the model has to be altered depending on the penalty term. A couple of years after Akaike's paper, Shibata [9] proposed another criterion, called RIC (Regularized Information Criterion), that took into account the penalty term, and allowed an optimal choice of the penalty parameter. This seemed like a perfect solution to Akaike's problem. However, the closed-form evaluation of RIC required the penalty term defined by a summation indexed in the same way as the (log) likelihood term. When this was not the case, Shibata [10] recommended the use of a bootstrap estimate of RIC. Unfortunately, this procedure required quite a bit of additional computations.

Akaike himself never produced user-friendly software for his proposed method, which nowadays is a must to popularize a statistical method. The high citation rate of his Psychometrika paper is probably due to the popularity of AIC in social science research in general, and not to the specific proposal he made in his paper. AIC has been proven useful in many situations other than determining the number of factors. For example, virtually all commercial software for structural equation modeling reports the value of AIC as one of the goodness-of-fit indices.

Akaike did not preclude the possibility of other information criteria. In fact, he originally used the acronym AIC to stand for "An Information Criterion," implying that there could be other criteria based on different rationales. Indeed, a host of other information criteria have subsequently been proposed, following Akaike's lead. Among those, perhaps the strongest competitor has been BIC (Bayesian Information Criterion) proposed by Schwarz [8], defined by

$$\mathrm{BIC}_\pi = -2\log(L_\pi) + (\log N)k, \tag{2}$$

where N is the sample size. BIC was derived as an approximation to (minus twice) the log posterior probability of model π under noninformative prior probabilities of the model [6]. As in AIC, the model corresponding to the smallest value of BIC is chosen as the best model. BIC tends to choose a more compact model than AIC because its penalty term is usually larger than that of AIC. As N increases, it is bound to choose the correct model with probability 1, if the correct model is included among the competing models. That is, BIC is consistent, while AIC has no such property. A large number of studies have been conducted to compare the relative performance of the two criteria (e.g., [12]). When applied to factor analysis with small to moderate sample sizes, BIC tends to underextract factors. Although the chance of missing a big factor is not high, if it happens, it can have serious detrimental effects on estimated parameters [5]. AIC, on the other hand, tends to overextract, although its effect is known to be relatively minor [4].

AIC's tendency to overextract is still a nuisance to be avoided as much as possible. Ogasawara [7] recently proposed a correction of AIC for its higher-order asymptotic bias. This statistic, called CAIC (Corrected AIC), is defined as

$$\mathrm{CAIC}_\pi = -2\log(L_\pi) + 2k + 4k^2/p(N-1), \tag{3}$$

where p is the number of observed variables. The additional term in CAIC tends to discourage overextraction of factors. A preliminary simulation study indicates that CAIC is more accurate than the original AIC in determining the correct number of factors.

References

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *The Second International Symposium on Information Theory*, (pp. 267–281). Budapest: Academiai Kiado. (Also in S. Kotz & N. L. Johnson (Eds.), *Breakthroughs in Statistics, Vol. 1. Foundations and Basic Theory*, (pp. 610–624). New York: Springer, 1992.)

[2] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.

[3] Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, *52*, 345–370.

[4] Fava, J. L., & Velicer, W. F. (1992). The effects of overextraction on factor and component analysis. *Multivariate Behavioral Research*, *27*, 387–415.

[5] Fava, J. L., & Velicer, W. F. (1996). The effects of underextraction in factor and component analysis. *Educational and Psychological Measurement*, *56*, 907–929.

[6] Neath, A. A., & Cavanaugh, J. E. (2011). The Bayesian information criterion: Background, derivation, and applications. *WIREs Computational Statistis 2012*, *4*, 199–203.

[7] Ogasawara, H. (in press). Bias correction of the Akaike information criterion in factor analysis. *Journal of Multivariate Analysis.*

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

[9] Shibata, R. (1989). Statistical aspects of model selection. In J. C. Willems (Ed.), *From Data to Models*, (pp. 501–522). Berlin: Springer.

[10] Shibata, R. (1997). Bootstrap estimate of Kullback-Leibler information for model selection. *Statistica Sinica*, *7*, 375–394.

[11] Takane, Y. (1994). A review of applications of AIC in psychometrics. In H. Bozdogan (Ed.), *Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modeling: An Informational Approach*, (pp. 379–403). Amsterdum: Kluver Academic Publishers.

[12] Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, *17*, 228–243.