On a Testing Procedure for Model Selection

Panaretos, John and Psarakis, Stelios

1991

Online at https://mpra.ub.uni-muenchen.de/6260/
MPRA Paper No. 6260, posted 13 Dec 2007 07:14 UTC
ON A TESTING PROCEDURE FOR MODEL SELECTION

J. Panaretos and S. Psarakis
University of Patras, Greece

ABSTRACT

In this paper a forecasting model selection scheme is considered which amounts to testing the predictive behaviour of a model by adopting Xekalaki and Katti's (1984) idea of assigning to its performance a score for each of a series of time points. The score reflects how close to, or how far from, the predictive value the observed actual value is. A statistical test is proposed for comparing the forecasting performances of two models.

Selecting the best of two competing models

Consider two linear models A and B of the form

\[ Y_t = X_t(M) \beta_t(M) + e_t(M) \]

where \( Y_t \) is an \( l \times 1 \) vector of observations on the dependent random variable, \( X_t(M) \) is an \( l \times m \) matrix of known coefficients, \( [X_t(M)X_t(M)] \neq 0 \), \( \beta_t(M) \) is an \( m \times 1 \) vector of regression coefficients, and \( e_t(M) \) is an \( l \times 1 \) vector of normal error random variables with \( \text{E}(e_t(M)) = 0 \) and \( \text{V}(e_t(M)) = \sigma^2 e_t(M) \). Here \( I_t \) is the \( l \times l \) identity matrix and \( M \) indexes the model (i.e., \( M = A \) or \( M = B \)). Therefore, a prediction for the value of the dependent random variable for time \( t+1 \) will be given by the statistic \( \hat{Y}^{(0)}_{t+1}(M) = X^{(0)}_{t+1}(M) \hat{\beta}_t(M) \), where \( \hat{\beta}_t(M) \) is the least squares estimator of \( \beta_t(M) \) at time \( t \). \( X^{(0)}_{t+1}(M) \) is a \( l \times m \) vector at time \( t+1 \). Let \( Y^{(0)}_{t+1} \) be the observed value of the dependent random variable at time \( t+1 \). Then

\[ \hat{Y}^{(0)}_{t+1}(M) - Y^{(0)}_{t+1} = e^{(0)}_{t+1}(M) \]

will follow the \( N(0, \sigma^2 M) \). (1)

and so [\( |\hat{Y}^{(0)}_{t+1}(M) - Y^{(0)}_{t+1}| \) will follow the folded normal distribution with mean \( \mu_f(M) = \sqrt{2/\pi} \sigma_M \) and variance \( \sigma_f^2(M) = \sigma^2 (1 - 2/\pi) \) (Leone et al. 1961). In other words

\[ \text{E}(|e^{(0)}(A)|) = \sqrt{2/\pi} \sigma_A \quad \text{and} \quad \text{E}(|e^{(0)}(B)|) = \sqrt{2/\pi} \sigma_B. \]

Suppose that we score the performance of model \( M \) by

\[ S(M) = \frac{1}{t} \sum_{i=1}^{t} \frac{|e^{(0)}_{t+1}(M)|}{\text{E}(e^{(0)}_{t+1}(M))}. \]
\[ |e_t(M)|: \text{Then, our selection will be based on the model with the minimum score. A natural choice of hypotheses to test could be:} \]

\[
H_0: E(|e_t(A)|) = E(|e_t(B)|) \\
H_1: E(|e_t(A)|) < E(|e_t(B)|)
\]  
(3)

Because of (2) this is equivalent to

\[
H_0: \sigma_A^2 = \sigma_B^2 \\
H_1: \sigma_A^2 < \sigma_B^2
\]  
(4)

Here we must note that

\[
\text{Cov}(|e_t(A)| + |e_t(B)|, |e_t(A)| - |e_t(B)|) = (\sigma_A^2 - \sigma_B^2)(1 - \frac{2}{n}).
\]  
(5)

Set \( R_t^+ = |e_t(A)| + |e_t(B)| \) and \( R_t^- = |e_t(A)| - |e_t(B)| \) Then, because of (5) the set of hypotheses (4), is equivalent to

\[
H_0: \text{Cov}(R_t^+, R_t^-) = 0 \\
H_1: \text{Cov}(R_t^+, R_t^-) < 0
\]  
or to

\[
H_0: \sigma_{R_t^+, R_t^-} = 0 \\
H_1: \sigma_{R_t^+, R_t^-} < 0.
\]

The obvious choice of a test statistic would be

\[
R = \frac{ \left[ \sum R_t^+ R_t^- / n - \overline{R_t^+} \overline{R_t^-} \right] }{ \sqrt{ \sum \left( R_t^+ - \overline{R_t^+} \right)^2 / n \sum \left( R_t^- - \overline{R_t^-} \right)^2 / n } }
\]

Then, under \( H_0 \) the asymptotic distribution of \( \sqrt{n} R \) is normal with mean zero and variance \( \text{Var}(R_t^+ R_t^-) / \text{Var}(R_t^+) \text{Var}(R_t^-) \) (Lehmann (1986)). So, values of \( \sqrt{n} R \) in the left tail of the normal distribution will call for rejection of \( H_0 \), thus indicating that model A performs better than model B.

**References**

