On a Testing Procedure for Model Selection

Panaretos, John and Psarakis, Stelios

1991

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

$$Y_t = X_t(M)\,\beta_M + e_t(M)$$

where Y_t is an $l_t \times 1$ vector of observations on the dependent random variable, $X_t(M)$ is an $l_t \times m_M$ matrix of known coefficients ($l_t \geq m_M$, $|X_t'(M)X_t(M)| \neq 0$), β_M is an $m_M \times 1$ vector of regression coefficients and $e_t(M)$ is an $l_t \times 1$ vector of normal error random variables with $E(e_t(M))=0$ and $V(e_t(M))=\sigma_M^2 I_t$ ($\sigma_M^2<\infty$). Here I_t is the $l_t \times l_t$ 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 β_M at time t, $X^0_{t+1}(M)$ is a $1 \times m_M$ vector at time t+1. Let Y^0_{t+1} be the observed value of the dependent random variable at time t+1. Then

$$\hat{Y}^0_{t+1}(M) - Y^0_{t+1} = e_{t+1}(M) \text{ will follow the } N(0,\sigma_M^2) \qquad (1)$$

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

$$E(|e_t(A)|)=\sqrt{2/\pi}\,\sigma_A \quad \text{and}$$
$$E(|e_t(B)|)=\sqrt{2/\pi}\,\sigma_B. \qquad (2)$$

Suppose that we score the performance of model M by

$|e_t(M)|$. Then, our selection will be based on the model with the minimum score. A natural choice of hypotheses to test could be:

$$H_0: E(|e_t(A)|) = E(|e_t(B)|) \qquad (3)$$
$$H_1: E(|e_t(A)|) < E(|e_t(B)|)$$

Because of (2) this is equivalent to

$$H_0: \sigma_A^2 = \sigma_B^2 \qquad (4)$$
$$H_1: \sigma_A^2 < \sigma_B^2$$

Here we must note that

$$\text{Cov.}(|e_t(A)|+|e_t(B)|, |e_t(A)|-|e_t(B)|) = (\sigma_A^2-\sigma_B^2)(1-\frac{2}{\pi}). \qquad (5)$$

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

$$H_0: \varrho_{R_{t+}R_{t-}}=0$$
$$H_1: \varrho_{R_{t+}R_{t-}}<0.$$

The obvious choice of a test statistic would be

$$R=\frac{\left[\frac{\sum R_{t+}R_{t-}}{n} - \overline{R_{t+}R_{t-}}\right]}{\sqrt{\frac{\sum\left(R_{t+}-\overline{R_{t+}}\right)^2}{n}\cdot\frac{\sum\left(R_{t-}-\overline{R_{t-}}\right)^2}{n}}},$$

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

Lehmann, E. L. (1986). "Testing Statistical Hypotheses". J.Wiley and Sons.

Leone, F.C. Nelson, L.S. and Nottigham, R.B. (1961). "The Folded Normal Distribution". Technometrics , 3(4), 543-550.

Xekalaki E. and Katti, S.K. (1984)."A Technique for Evaluating Forecasting models". Biom. J., 26, 173-184.