On a Distribution Arising in the Context of Comparative Model Performance Evaluation Problems

Panaretos, John and Psarakis, Stelios and Xekalaki, Evdokia

1998

ON A DISTIRIBUTION ARISING IN THE CONTEXT OF COMPARATIVE MODEL PERFORMANCE EVALUATION PROBLEMS

J. Panaretos S. Psarakis & E. Xekalaki
Department of Statistics
Athens University of Economics & Business
76, Patision Str. 10434
Athens, GREECE
e-mail: jpan@aueb.gr

Key words and phrases: model selection, bivariate gamma distribution, F distribution.

ABSTRACT

The paper deals with a distribution that arises as the distribution of a sample statistic used to compare the predictive ability of two competing linear models. It is defined as the distribution of the ratio of two correlated gamma variables and its probabilities are tabulated in order that they become readily available for practical use.

INTRODUCTION

Evaluating the forecasting potential of a model before this can be used for planning and decision making has been the concern of many statistical workers. A number of evaluation techniques has thus been considered and much theory has been developed.

Xekalaki and Katti (1984) introduced an evaluation scheme of sequential nature based on the idea of scoring rules in which the researcher's subjectivity plays an important role. Its effect is reflected through the rules according to which the performance of the model is scored and rated.

In this paper an evaluation method is suggested which is again based on the use of a scoring rule but is free of the element of subjectivity. A scoring rule is suggested to rate the behaviour of a linear forecasting model for each of a series of n points in time. A final rating which embodies the step by step scores is then used as a test statistic for testing the adequacy of the model.

The problem of comparative evaluation is also considered and a hypothesis test is suggested in the case of two competing linear models. In this case a new distribution, which is a generalized form of the F distribution, is considered. This distribution and the scoring rule it generates are used to compare two linear models used on real data.

COMPARATIVE EVALUATION OF THE PREDICTIVE ABILITY OF TWO LINEAR MODELS WITH THE USE OF A GENERALISED FORM OF THE F DISTRIBUTION

Consider the linear model

$$\mathbf{Y}_t = \mathbf{X}_t\boldsymbol{\beta} + \mathbf{e}_t \qquad (1)$$

where $\mathbf{Y}_t$ is an $\ell_t \times 1$ vector of observations on the dependent random variable, $\mathbf{X}_t$ is an $\ell_t \times m$ matrix of known coefficients $\left(\ell_t \geq m, \left|\mathbf{X}_t'\mathbf{X}_t\right| \neq 0\right)$, $\boldsymbol{\beta}$ is an $m \times 1$ vector of regression coefficients and $\mathbf{e}_t$ is an $l \times 1$ vector of normal error random variables with $E(\mathbf{e}_t)=0$ and $V(\mathbf{e}_t)=\sigma^2 I_t$. Here I_t is the $\ell_t \times \ell_t$ identity matrix. Therefore, a prediction for the value of the dependent random variable for time t+1 will be given by the statistic

$$\hat{\mathbf{Y}}_{t+1}^{0} = \mathbf{X}_{t+1}^{0}\hat{\boldsymbol{\beta}}_t$$

where $\hat{\boldsymbol{\beta}}_t$ is the least squares estimator of $\boldsymbol{\beta}$ at time t and $\mathbf{X}_{t+1}^{0}$ is a $1 \times m$ vector of values of the regressors at time t+1, t=0,1,2,...
Obviously

$$\mathbf{X}_{t+1} = \begin{bmatrix} \mathbf{X}_t \\ \mathbf{X}_{t+1}^{0} \end{bmatrix} \qquad \text{and} \qquad \mathbf{Y}_{t+1} = \begin{bmatrix} \mathbf{Y}_t \\ \mathbf{Y}_{t+1}^{0} \end{bmatrix}$$

The predictive behaviour of the model would naturally be evaluated by a measure that would be based on a statistic reflecting the degree of agreement of the observed actual value $\mathbf{Y}_{t+1}^{0}$ to the predicted value $\hat{\mathbf{Y}}_{t+1}^{0}$. Such a statistic may be $\left|\mathbf{r}_{t+1}\right|$, where

$$\mathbf{r}_{t+1} = \frac{\hat{\mathbf{Y}}_{t+1}^{0} - \mathbf{Y}_{t+1}^{0}}{S_t\sqrt{\left(1 + \mathbf{X}_{t+1}^{0}\left(\mathbf{X}_t'\mathbf{X}_t\right)^{-1}\mathbf{X}_{t+1}'^{0}\right)}} \qquad t = 0,1,2,\ldots \quad (2)$$

Obviously, $\left|\mathbf{r}_{t+1}\right|$ is merely an estimate of the standardized distance between the predicted and the observed value of the dependent random variable when σ^2 is estimated on the basis of the preceding ℓ_t observations available at time t. S_t^2 is given by

$$S_t^2 = \frac{\left[\mathbf{Y}_t - \mathbf{X}_t\hat{\boldsymbol{\beta}}_t\right]'\left[\mathbf{Y}_t - \mathbf{X}_t\hat{\boldsymbol{\beta}}_t\right]}{\left[\ell_t - m\right]}, \qquad t = 0,1,2,\ldots.$$

So, a score based on $\left|\mathbf{r}_{t+1}\right|$ can provide a measure of the adequacy of the model for each of a series of n points in time. Then, as a final rating of the model one can consider the average of these scores or any other summary statistic that can be regarded as reflecting the forecasting potential of the model.

In the sequel, we consider using $\mathbf{r}_t^2$ as a scoring rule to rate the performance of the model at time t for a series of n points in time, (t=1,2,...n) and we define

$$R_n = \sum_{t=1}^{n} r_t^2 / n \qquad (3)$$

the average of the squared recursive residuals, to be the final rating of the model.

It has been shown (Brown et. al. 1975, Kendall et.al. 1983)) that if $\mathbf{e}_t$ is a vector of normal error variables with $E(\mathbf{e}_t)=0$ and $V(\mathbf{e}_t)=\sigma^2 I_t$, the quantities

$$w_{t+1} = \frac{\hat{Y}^0_{t+1} - Y^0_{t+1}}{\left(1 + \mathbf{X}^0_{t+1}\left(\mathbf{X}_t'\mathbf{X}_t\right)^{-1}\mathbf{X}^0_{t+1}\right)} \qquad t=0,1,2,...$$

are independently and identically distributed normal variables with mean 0 and variance σ^2. Then, according to Kotlarski's (1966) characterization of the normal distribution by the t distribution, the quantities $r_{t+1} = w_{t+1}/s_t$ t=0,1,2,... constitute a sequence of independent t variables with ℓ_t -m degrees of freedom, t=0,1,2,... Hence, by the assumptions of the model considered, and for large ℓ_0, the variables r_{t+1}, t=0,1,2,... as given by (2) constitute a sequence of approximately standard normal variables which are mutually independent. This implies that

$$n R_n = \sum_{t=1}^{n} r_t^2$$

is a chi-square variable with n degrees of freedom. (Psarakis and Panaretos (1990)).

Let now A and B be two competing linear models in the sense of (1) that have been used for prediction purposes for a number n_1 and n_2 of years respectively. Then, a decision on whether model A is more adequate than model B, would naturally be based on the ratio of

$$R_{n_1,n_2} = \frac{R_{n_1}(A)}{R_{n_2}(B)} \qquad (4)$$

where $R_{n_1}(A)$, $R_{n_2}(B)$ are given by (2.3) for $n=n_1$ and $n=n_2$, respectively.

For large ℓ_{t_1}, ℓ_{t_2}, the statistic R_{n_1,n_2} will be an F variable with n_1 and n_2 degrees of freedom whenever the ratings of the two models are independent. Hence, values of R_{n_1,n_2} in the right tail of the F distribution with n_1 and n_2 degrees of freedom will indicate a higher performance by model A. However under the conditions of the problem under consideration the assumption of independence does not seem to be satisfied. Although Kotlarski (1964) has shown that, under certain conditions, the quotient X/Y follows the F distribution even when X and Y are dependent variables Panaretos, Psarakis & Xekalaki (1997) proved that Kotlarski's conditions for this to happen are not met in our problem.

It has been shown (Panaretos, Psarakis & Xekalaki (1997)) that the analytic form of the distribution of the quotient X/Y is :

$$f_{X/Y}(z) = \frac{\left(1-\rho^2\right)^k}{B(k,k)} z^{k-1}(1+z)^{-2k}\left[1-\left[\frac{2\rho}{z+1}\right]^2 z\right]^{-\frac{2k+1}{2}} \quad (5)$$

We refer to the above distribution as the **correlated gamma-ratio distribution.** In the case where X and Y are independent (i.e. when ρ=0) the probability density function of the quotient X/Y will follow the F distribution with 2k and 2k degrees of freedom. Percentage points for different values of k and ρ have been calculated and presented in Panaretos, Psarakis & Xekalaki (1997). These tables can be very handy in testing hypotheses concerning the adequacy of two competing models.

For the purpose of illustrating the model selection procedure, a problem of selecting a linear model by the USDA, presented in Xekalaki and Katti (1984) to predict the soybean yield in the State of Iowa for the years 1956 to 1980 was re-examined based on several sets of real data. So, in testing the hypotheses

H_0: *models A and B are as adequate*

H_1: *A is better than B* ,

it was found that at the 5% significance level there was enough evidence to reject the null hypotheses for two crop reporting districts while there was not enough evidence to reject the null hypotheses for the rest of the crop reporting districts considered.

REFERENCES

Brown, R. L., Durbin, J. and Evans, J. M. (1975) Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society B*, **24**, 149-192.

Kendall, M. Stuart, A. and Ord, J. K. (1983) The Advanced Theory of Statistics. Vol .1 (4th edition). Griffin London.

Kibble, W. F. (1941) A Two Variate Gamma Type Distribution. *Sankhya*, **5**, 137-150.

Kotlarski, I. (1964) On bivariate random variables where the quotient of their coordinates follows some known distributions. *Annals of Mathematical Statistics*, 1673-1684.

Kotlarski, I. (1966) On characterizing the normal distribution by the Student's law. *Biometrika*, **53**, 603-606.

Panaretos, J., Psarakis S., & Xekalaki E. (1997) The correlated Gamma-ratio distribution in model evaluation and selection. Technical Report 33, Department of Statistics, Athens University of Economics and Business, November 1997.

Psarakis, S. & Panaretos, J. (1990) On an approach to model evaluation and selection. XV Biometric Conference. Budapest, Hungary, July 1990.

Xekalaki, E. & Katti, S. K. (1984) A Technique for Evaluating Forecasting Models. *Biometrical Journal*, **26(1)**, 173-184.