On non-nested regression models

Jiří Anděl

Abstract. A generalization of a test for non-nested models in linear regression is derived for the case when there are several regression models with more regressors.

Keywords: non-nested models, regression analysis

Classification: 62J05

1. Introduction.

Consider a regression model

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + e_i, \qquad i = 1, \dots, n, \tag{1.1}$$

where $e_1, \dots, e_n$ are i.i.d. $N(0, \sigma^2)$ random variables with an unknown variance $\sigma^2 > 0$. Let S_e be residual sum of squares (RSS) in this model. If the matrix

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_1 & z_1 \\ \dots & \dots & \dots \\ 1 & x_n & z_n \end{pmatrix}$$

has rank $r = 3$ then it is easy to test if the model (1.1) is significantly better than the model

$$Y_i = \beta_0 + \beta_1 x_i + e_i, \qquad i = 1, \dots, n. \tag{1.2}$$

It suffices to test the hypothesis $H_0 : \beta_2 = 0$ against $H_1 : \beta_2 \neq 0$, which is an elementary procedure described in statistical textbooks. However, the problem which of the models (1.2) and

$$Y_i = \beta_0 + \beta_2 z_i + e_i, \qquad i = 1, \dots, n, \tag{1.3}$$

is significantly better, is more complicated. This problem is very important in applications. For example, choosing $z_i = \ln x_i$ we can ask if the model $Y_i = \beta_0 + \beta_2 \ln x_i + e_i$ is better than the model $Y_i = \beta_0 + \beta_1 x_i + e_i$ or not. It is clear that such decision can play an important role especially in statistical analysis of biological and econometrical data.

The models (1.2) and (1.3) are called non-nested or separate.

A method for comparing the models (1.2) and (1.3) was published by Hotelling (1940). His motivation was to test whether the correlation coefficient between Y and

x is significantly different from the correlation coefficient between Y and z. Healy (1955) showed that Hotelling's procedure is equivalent to a test about regression coefficients. This idea was generalized to a larger number of models of the type (1.2) by Williams (1959), who also pointed out that Healy's result is not correct. In a note which is published in Williams' paper Healy apologizes for the error.

The following simple description of the method for comparing (1.2) and (1.3) is taken from Kendall and Stuart (1967), Exercise 28.22.

It is well known that

$$\mathrm{RSS}_1 = \sum (Y_i - \bar{Y})^2 - \left[\sum (x_i - \bar{x})(Y_i - \bar{Y})\right]^2 \Big/ \sum (x_i - \bar{x})^2$$

and

$$\mathrm{RSS}_2 = \sum (Y_i - \bar{Y})^2 - \left[\sum (z_i - \bar{z})(Y_i - \bar{Y})\right]^2 \Big/ \sum (z_i - \bar{z})^2$$

are residual sums of squares in the models (1.2) and (1.3), respectively. Let r be the sample correlation coefficient between x_i and z_i, $i = 1, \dots, n$. Define

$$u_i = \frac{x_i - \bar{x}}{\sqrt{\sum (x_i - \bar{x})^2}}, \qquad v_i = \frac{z_i - \bar{z}}{\sqrt{\sum (z_i - \bar{z})^2}},$$

$$U = \sum Y_i u_i, \qquad V = \sum Y_i v_i.$$

It can be easily checked that

$$var\ U = var\ V = \sigma^2, \qquad cov(U, V) = \sigma^2 r, \qquad var(U - V) = 2\sigma^2(1 - r).$$

It is clear that

$$\mathrm{RSS}_1 = \sum (Y_i - \bar{Y})^2 - U^2, \qquad \mathrm{RSS}_2 = \sum (Y_i - \bar{Y})^2 - V^2. \tag{1.4}$$

If U is not significantly different from V then also RSS_1 is not significantly different from RSS_2. If $EU = EV$ then

$$U - V \sim N[0, 2\sigma^2(1 - r)].$$

An unbiased estimator for σ^2 in the model (1.1) is $s^2 = S_e/(n-3)$. Since s^2 is independent of (U, V), under $H_0 : EU = EV$ the statistic

$$T = \frac{U - V}{\sqrt{2s^2(1 - r)}}$$

has the t_{n-3} distribution. If $|T| \geq t_{n-3}(\alpha)$, where $t_{n-3}(\alpha)$ is the critical value, we reject H_0 .

Notice, however, that for comparison of the models (1.2) and (1.3) we should rather test the hypothesis $H_0^*: E\mathrm{RSS}_1 = E\mathrm{RSS}_2$, i.e. that $EU^2 = EV^2$ instead of H_0 mentioned above. This is a drawback of the mentioned method.

A generalization of the described procedure is introduced in Section 2.

A different approach used for analysis of non-nested models was proposed by Cox (1962). It is an extension of the likelihood ratio test. The theory of testing separate models is a growing area with many applications. The most popular tests are

(1) the orthodox F-test;
(2) the J-test (see Davidson and MacKinnon 1981);
(3) the JA-test (see Fisher and McAleer 1981).

More detailed information can be found in the review articles by MacKinnon (1983) and McAleer (1987). The book by Doran (1989), Chapter 14.5, can be recommended as a good elementary introduction to such problems.

2. Several regression models with more regressors.

Consider a regression model

$$Y_i = \beta_0' + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + e_i, \qquad i = 1, \ldots, n, \tag{2.1}$$

where $e_1, \ldots, e_n$ are i.i.d. $N(0, \sigma^2)$ random variables and the matrix

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1k} \\ \ldots & \ldots & \ldots & \ldots \\ 1 & x_{n1} & \ldots & x_{nk} \end{pmatrix}$$

has rank $k+1$. Let

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij}, \qquad j = 1, \ldots, k.$$

The model (2.1) can be equivalently written in the form

$$Y_i = \beta_0 + \beta_1 (x_{i1} - \bar{x}_1) + \cdots + \beta_k (x_{ik} - \bar{x}_k) + e_i, \qquad i = 1, \ldots, n, \tag{2.2}$$

where

$$\beta_0 = \beta_0' - \beta_1 \bar{x}_1 - \cdots - \beta_k \bar{x}_k.$$

The matrix form of (2.2) is

$$\boldsymbol{Y} = (\boldsymbol{1}, \boldsymbol{H})\boldsymbol{\beta} + \boldsymbol{e} \tag{2.3}$$

where

$$\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \ldots \\ Y_n \end{pmatrix}, \qquad \boldsymbol{1} = \begin{pmatrix} 1 \\ \ldots \\ 1 \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \ldots \\ \beta_k \end{pmatrix}, \qquad \boldsymbol{e} = \begin{pmatrix} e_1 \\ \ldots \\ e_n \end{pmatrix},$$

$$\boldsymbol{H} = \begin{pmatrix} x_{11} - \bar{x}_1, & \ldots, & x_{1k} - \bar{x}_k \\ \ldots & \ldots & \ldots \\ x_{n1} - \bar{x}_1, & \ldots, & x_{nk} - \bar{x}_k \end{pmatrix}.$$

The residual sum of squares in the model (2.3) is

$$S_e = \boldsymbol{Y}'\boldsymbol{Y} - n\bar{Y}^2 - \boldsymbol{Y}'\boldsymbol{H}(\boldsymbol{H}'\boldsymbol{H})^{-1}\boldsymbol{H}'\boldsymbol{Y}$$

and the least squares estimators for β_0 and $(\beta_1, \dots, \beta_k)'$ are $\bar{Y}$ and $(\boldsymbol{H}'\boldsymbol{H})^{-1}\boldsymbol{H}'\boldsymbol{Y}$, respectively. These estimators are independent of S_e.

Now, consider the submodels

$$\boldsymbol{Y} = (\boldsymbol{1}, \boldsymbol{H}_i)\boldsymbol{\alpha} + \boldsymbol{e}, \qquad i = 1, \dots, m \tag{2.4}$$

where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_c)'$ and where each matrix $\boldsymbol{H}_i$ consists of some c columns of the matrix $\boldsymbol{H}$. The residual sum of squares RSS_i of the i-th model (2.4) is

$$\mathrm{RSS}_i = \boldsymbol{Y}'\boldsymbol{Y} - n\bar{Y}^2 - \boldsymbol{Y}'\boldsymbol{H}_i(\boldsymbol{H}_i'\boldsymbol{H}_i)^{-1}\boldsymbol{H}_i'\boldsymbol{Y}.$$

Define

$$\boldsymbol{U}_i = (\boldsymbol{H}_i'\boldsymbol{H}_i)^{-1/2}\boldsymbol{H}_i'\boldsymbol{Y}, \qquad i = 1, \dots, m.$$

We have

$$\mathrm{RSS}_i = \boldsymbol{Y}'\boldsymbol{Y} - n\bar{Y}^2 - \boldsymbol{U}_i'\boldsymbol{U}_i.$$

If $\boldsymbol{U}_1, \dots, \boldsymbol{U}_m$ do not differ substantially then also $\mathrm{RSS}_1, \dots, \mathrm{RSS}_m$ do not differ very much and all the models (2.4) can be considered as equally successful (or equally unsuccessful). A test which enables us to decide if $\boldsymbol{U}_1, \dots, \boldsymbol{U}_m$ are significantly different can be based on the following theorem.

Theorem 2.1. *Define*

$$\boldsymbol{F}_i = (\boldsymbol{H}_i'\boldsymbol{H}_i)^{-1/2}\boldsymbol{H}_i', \qquad \boldsymbol{V}_{ij} = \boldsymbol{F}_i\boldsymbol{F}_j' \quad \text{for } i, j = 1, \dots, m, \qquad \boldsymbol{V} = (\boldsymbol{V}_{ij})_{i,j=1}^m.$$

Let the matrix $\boldsymbol{V}$ be regular. Denote by $\boldsymbol{V}^{ij}$ the $c \times c$ blocks of the matrix $\boldsymbol{V}^{-1}$ such that $\boldsymbol{V}^{-1} = (\boldsymbol{V}^{ij})_{i,j=1}^m$. Define

$$\boldsymbol{u} = \left(\sum_i \sum_j \boldsymbol{V}^{ij}\right)^{-1} \sum_i \sum_j \boldsymbol{V}^{ij}\boldsymbol{U}_j.$$

Let $s^2 = S_e/(n-k-1)$ be an estimator of σ^2 in the model (2.3) with $n-k-1$ degrees of freedom. If $E\boldsymbol{U}_1 = \dots = E\boldsymbol{U}_m$ then

$$Z = \frac{1}{c(m-1)s^2} \sum_i \sum_j (\boldsymbol{U}_i - \boldsymbol{u})'\boldsymbol{V}^{ij}(\boldsymbol{U}_j - \boldsymbol{u})$$

has the F-distribution with $c(m-1)$ and $n-k-1$ degrees of freedom.

Proof: First of all we prove that the matrix $\sum\sum \boldsymbol{V}^{ij}$ is regular. Let $\boldsymbol{I}$ be the $c \times c$ unit matrix and define a $c \times cm$ matrix $\boldsymbol{K} = (\boldsymbol{I}, \dots, \boldsymbol{I})$. We have

$$\sum\sum \boldsymbol{V}^{ij} = \boldsymbol{K}\boldsymbol{V}^{-1}\boldsymbol{K}' = (\boldsymbol{K}\boldsymbol{V}^{-1/2})(\boldsymbol{K}\boldsymbol{V}^{-1/2})'.$$

The rank of $\boldsymbol{K}$ is c, $\boldsymbol{V}^{-1/2}$ is regular and thus the rank of $\boldsymbol{K}\boldsymbol{V}^{-1/2}$ is also c. Since the rank of a matrix $\boldsymbol{G}$ is equal to the rank of $\boldsymbol{G}\boldsymbol{G}'$, the matrix $\sum\sum \boldsymbol{V}^{ij}$ of the type $c \times c$ has also rank c.

It is easy to check that $var(\boldsymbol{U}_1', \dots, \boldsymbol{U}_m')' = \sigma^2 \boldsymbol{V}$. Define

$$Z^* = \sigma^{-2} \sum\sum (\boldsymbol{U}_i - \boldsymbol{u})' \boldsymbol{V}^{ij} (\boldsymbol{U}_j - \boldsymbol{u}).$$

After a computation we get

$$Z^* = \sigma^{-2} \sum_i \sum_j \boldsymbol{U}_i' \left[\boldsymbol{V}^{ij} - \sum_t \boldsymbol{V}^{it} \left(\sum_\alpha \sum_\beta \boldsymbol{V}^{\alpha\beta} \right)^{-1} \sum_w \boldsymbol{V}^{wj} \right] \boldsymbol{U}_j.$$

Let $\boldsymbol{A}$ be the matrix with $c \times c$ blocks

$$\boldsymbol{A}_{ij} = \boldsymbol{V}^{ij} - \sum_t \boldsymbol{V}^{it} \left(\sum_\alpha \sum_\beta \boldsymbol{V}^{\alpha\beta} \right)^{-1} \sum_w \boldsymbol{V}^{wj}.$$

It can be verified directly that the matrix $\boldsymbol{A}\boldsymbol{V}$ is idempotent and that its trace is $c(m-1)$. It implies that the rank of $\boldsymbol{A}\boldsymbol{V}$ is also $c(m-1)$. The variable Z^* does not depend on the value $EX_1 = \dots = EX_n$. Without loss of generality we can assume in this proof that $EX_1 = 0$. Corollary 2.2 in Searle (1971), p. 58, implies that Z^* has the χ^2-distribution with $c(m-1)$ degrees of freedom. Since $\boldsymbol{U}_i$ depends on $\boldsymbol{Y}$ only through $\boldsymbol{H}_i\boldsymbol{Y}$, we can see that $(\boldsymbol{U}_1, \dots, \boldsymbol{U}_m)$ and S_e are independent. But S_e/σ^2 has the χ^2-distribution with $n-k-1$ degrees of freedom and thus Z has the $F_{c(m-1),n-k-1}$-distribution. □

Theorem 2.2. *The matrix* $\boldsymbol{V}$ *in Theorem* 2.1 *is regular if and only if all the columns of the matrix*

$$\boldsymbol{G} = (\boldsymbol{H}_1, \dots, \boldsymbol{H}_m)$$

are different.

Proof: Define

$$\boldsymbol{L} = \begin{pmatrix} (\boldsymbol{H}_1'\boldsymbol{H}_1)^{-1/2} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \dots & \dots & \dots & \dots \\ \boldsymbol{0} & \boldsymbol{0} & \dots & (\boldsymbol{H}_m'\boldsymbol{H}_m)^{-1/2} \end{pmatrix}.$$

It can be easily checked that

$$\boldsymbol{V} = \boldsymbol{L}\boldsymbol{G}'\boldsymbol{G}\boldsymbol{L}.$$

Let $r(\boldsymbol{A})$ denote the rank of a matrix $\boldsymbol{A}$. Since $\boldsymbol{L}$ is regular and $r(\boldsymbol{G}'\boldsymbol{G}) = r(\boldsymbol{G})$, we have $r(\boldsymbol{V}) = r(\boldsymbol{G})$. But all the columns of the matrix $\boldsymbol{G}$ are columns of the matrix $\boldsymbol{H}$, which is supposed to have linearly independent columns. □

Thus $\boldsymbol{V}$ is regular if and only if no two matrices $\boldsymbol{H}_i$, $\boldsymbol{H}_j$ $(i \neq j)$ contain the same column of the original matrix $\boldsymbol{H}$.

Let us remark that $\boldsymbol{u}$ is the best linear unbiased estimator of the common expectation $E\boldsymbol{U}_1 = \dots = E\boldsymbol{U}_m$.

The hypothesis that all the submodels (2.4) are equally suitable for description of $\boldsymbol{Y}$ is rejected when the variable Z defined in Theorem 2.1 exceeds the critical value $F_{c(m-1),n-k-1}(\alpha)$.

References

Cox D.R., *Further results on test of separate families of hypotheses*, J. Roy. Statist. Soc. Ser. B **24** (1962), 406–424.

Davidson R., MacKinnon J.G., *Some non-nested hypothesis tests and the relations among them*, Rev. Econom. Stud. **49** (1982), 551–565.

Doran H.E., *Applied Regression Analysis in Econometrics*, Dekker, New York and Basel, 1989.

Fisher G., McAleer M., *Alternative procedures and associated tests of significance for non-nested hypotheses*, J. Econometrics **16** (1981), 103–119.

Hotelling H., *The selection of variates for use in prediction, with some comments on the general problem of nuisance parameters*, Ann. Math. Statist. **11** (1940), 271–283.

Healy M.J.R., *A significance test for the difference in efficiency between two predictors*, J. Roy. Statist. Soc. Ser. B **17** (1955), 266–268.

Kendall M.G., Stuart A., *The Advanced Theory of Statistics. Vol. 2: Inference and Relationship*, Griffin, London, second ed.

MacKinnon J.G., *Model specification tests against non-nested alternatives*, Econometric Rev. **2** (1983), 85–110.

McAleer M., *Specification tests for separate models: a survey*, In Specification Analysis in the Linear Model (King M.L. and Giles D.E.A., eds.), Routledge & Kegan Paul, London, 1987.

Searle S.R., *Linear Models*, Wiley, New York, 1971.

Williams E.J., *The comparison of regression variables*, J. Roy. Statist. Soc. Ser. B **21** (1959), 396–399.

Department of Statistics, Faculty of Mathematics and Physics, Sokolovská 83, 186 00 Praha 8, Czech Republic

(*Received* September 8, 1992, *revised* December 12, 1992)