TESTS OF EQUALITY BETWEEN SETS OF COEFFICIENTS IN TWO LINEAR REGRESSIONS

BY GREGORY C. CHOW

Having estimated a linear regression with \( p \) coefficients, one may wish to test whether \( m \) additional observations belong to the same regression. This paper presents systematically the tests involved, relates the prediction interval (for \( m = 1 \)) and the analysis of covariance (for \( m > p \)) within the framework of general linear hypothesis (for any \( m \)), and extends the results to testing the equality between subsets of coefficients.

1. INTRODUCTION

The model of normal linear regression has often been widely applied to the measurement of economic relationships. In studies of the consumption function, the mean of consumption is assumed to be a linear function of income and other variables. In studies of consumer demand, the quantity of a commodity is regressed linearly on its price, income, and perhaps the price of an important complement or substitute. In studies of business investment, linear regressions on profits, sales, liquid asset holdings, and the interest rate, have been estimated. Other notable examples include empirical studies of dividend policy, of prices of corporate stocks, and of cost and supply functions.

When a linear regression is used to represent an economic relationship, the question often arises as to whether the relationship remains stable in two periods of time, or whether the same relationship holds for two different groups of economic units. Is the consumption pattern of the American people today the same as it was before World War II? Do the firms in the steel industry and the firms in the chemical industry have similar dividend policies? Statistically these questions can be answered by testing whether two sets of observations can be regarded as belonging to the same regression model.

Often there is no economic rationale in assuming that two relationships are completely the same. It may be more reasonable to suppose that only parts of the relationships are identical in two periods, or for two groups. Maybe the price elasticity of demand for a certain food product has not changed since World War II, while the income elasticity has changed. Maybe the investments of two groups of firms are affected in the same manner by profits, but not by liquid assets. Statistically, we are asking whether subsets of coefficients in two regressions are equal.

1 An early draft of this paper has been revised after helpful comments from William Kruskal, Edwin Kuh, and David L. Wallace, to all of whom I am grateful.
To state our problems more formally, let \( y \) be the dependent variable, and \( x_1, x_2, \ldots, x_p \) be the explanatory variables. Assume that there is a sample of \( n \) observations. These observations are governed by a model of normal linear regression. In matrix notations, the model is:

\[
\begin{bmatrix}
y_1 \\
y_2 \\
\vdots \\
y_n
\end{bmatrix} =
\begin{bmatrix}
x_{11} & x_{12} & \cdots & x_{1p} \\
x_{21} & x_{22} & \cdots & x_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
x_{n1} & x_{n2} & \cdots & x_{np}
\end{bmatrix}
\begin{bmatrix}
\beta_1 \\
\beta_2 \\
\vdots \\
\beta_p
\end{bmatrix}
+ 
\begin{bmatrix}
\epsilon_1 \\
\epsilon_2 \\
\vdots \\
\epsilon_n
\end{bmatrix}
\]

Here the \( x \)'s are \( p \) fixed variates. The \( \beta \)'s are the regression coefficients—\( \beta_1 \) is the intercept if \( x_1 \) is set identically equal to one. The \( \epsilon \)'s are independent and normally distributed, each with mean zero and standard deviation \( \sigma \). Assuming \( n > p \) and nonsingularity of the \( X \) matrix, we can estimate the parameters \( \beta_1, \beta_2, \ldots, \beta_p \) and \( \sigma \). Our problems are the testing of whether \( m \) additional observations are from the same regression as the first sample of \( n \) observations, and the testing of whether subsets of coefficients in the two regressions are identical. The present paper is devoted to a systematic and unified treatment of these tests.

To test the hypothesis that both samples belong to the same regression, the well-known prediction interval [8] can be used when the number \( m \) of observations in the second sample equals one, and the analysis of covariance [7] can be used when \( m > p \). We will present two tests for the case \( 2 \leq m \leq p \). The first test, to be presented in Section 2, is based on a prediction interval for the mean of \( m \) additional observations. The second test is an \( F \) test, to be developed in Section 3. The relationship among this \( F \) test, the prediction interval, and the analysis of covariance will be explained in Section 4. In Section 5, our results will be extended to testing the equality between subsets of regression coefficients in the two regressions. Two examples of econometric applications are given in Section 6. These examples are concerned with the temporal stability of a statistical demand function for automobile ownership, and of a statistical demand function for new automobiles. It was through these examples that I became interested in the tests presented in this paper.

2. PREDICTION INTERVAL FOR THE MEAN OF \( m \) ADDITIONAL OBSERVATIONS

It is straightforward to extend the prediction interval idea from one observation to the arithmetic mean of \( m \) observations.

First, let us rewrite the model in Section 1 briefly as

\[
y_1 = X_1 \beta_1 + \epsilon_1
\]
where both $\gamma_1$ and $\epsilon_1$ are column vectors with $n$ elements, $X_1$ is a nonsingular $n$ by $p$ matrix, and $\beta_1$ is the column vector of the $p$ regression coefficients. The subscript 1 denotes the first sample of $n$ observations. The least-squares estimator of $\beta_1$ from this first sample is given by:

$$b_1 = (X_1^tX_1)^{-1}X_1\gamma_1 = \beta_1 + (X_1^tX_1)^{-1}X_1\epsilon_1$$

$X_1^tX_1$ is the cross-product matrix of the $p$ $x$'s from the first sample.

Let the $m$ additional observations $\gamma_2$ of the dependent variable be specified by the model

$$\gamma_2 = X_2\beta_2 + \epsilon_2.$$  

$X_2$ is a nonsingular $m$ by $p$ matrix, with its $m$ rows representing the $m$ new observations on the $p$ explanatory variables. $\epsilon_2$ is normally distributed with the covariance matrix $I\sigma^2$. If we form the difference between the vector $\gamma_2$ and the vector of predictions based on the regression estimated by the first $n$ observations, we have, incorporating the relations (2) and (3),

$$d = \gamma_2 - X_2b_1 = X_2\beta_2 - X_2\beta_1 + \epsilon_2 - X_2(X_1^tX_1)^{-1}X_1\epsilon_1.$$

The expectation of $d$ is

$$E(d) = X_2\beta_2 - X_2\beta_1.$$  

Because of the independence of $\epsilon_2$ and $\epsilon_1$, the covariance matrix of $d$ becomes

$$\text{Cov} (d) = \text{Cov} (\epsilon_2) + \text{Cov} [X_2(X_1^tX_1)^{-1}X_1^t\epsilon_1]$$

$$= I\sigma^2 + X_2(X_1^tX_1)^{-1}X_1^t(Cov \epsilon_1)X_1(X_1^tX_1)^{-1}X_2^t$$

$$= [I + X_2(X_1^tX_1)^{-1}X_2^t]\sigma^2.$$  

In the special case when $m = 1$, both $\gamma_2$ and $d$ become scalars, and $X_2$ becomes a row vector. From (6), the variance of $d$ in this special case will be

$$\text{Var} (d) = [1 + X_2(X_1^tX_1)^{-1}X_2^t]\sigma^2$$

$\sigma^2$ can be estimated by $s^2$, the (unbiased) square of the standard error from the first $n$ observations. Under the null hypothesis that $\beta_2 = \beta_1 = \beta$ the expectation of $d$ given in (5) will be zero, and the ratio

$$\frac{d^2}{[1 + X_2(X_1^tX_1)^{-1}X_2^t]s^2}$$

will be distributed as $F(1, n - p)$. This test, which is based on the prediction interval for one new observation, can be found in [8].
When we have \( m \) new observations and thus \( m \) differences \( d_1, d_2, \ldots, d_m \), we may consider the average.

\[
\bar{d} = \frac{1}{m} \sum_{i=1}^{m} d_i
\]

Given the covariance matrix of \( d \) in (6) above, the variance of \( \bar{d} \) is

\[
\text{Var} (\bar{d}) = \frac{1}{m^2} \text{Var} \left[ \sum_{i=1}^{m} d_i \right] = \frac{s^2}{m^2} \left[ \begin{bmatrix} 1 & \ldots & 1 \end{bmatrix} [I + X_2(X_1'X_1)^{-1} X_2'] \begin{bmatrix} 1 \\ \vdots \end{bmatrix} \right]
\]

Similarly, under the null hypothesis that \( \beta_2 = \beta_1 \),

\[
\frac{\bar{d}^2}{\left\{ \begin{bmatrix} 1 & \ldots & 1 \end{bmatrix} [I + X_2(X_1'X_1)^{-1} X_2'] \begin{bmatrix} 1 \\ \vdots \end{bmatrix} \right\} \frac{s^2}{m^2}}
\]

will be distributed as \( F(1, m - \hat{p}) \).

The use of a \( \bar{d} \) test can be found in reference [4], although the formula given in the Appendix of this reference is incorrect.\(^2\) There is little rationale in using \( \bar{d} \) for the purpose of testing. The test is obviously weak against a number of alternative hypothesis. One can envisage many situations in which \( \bar{d} \) is small, not because the new \( m \) observations have come from the same regression, but because their deviations cancel out. The usefulness of deriving the distribution of \( \bar{d} \) lies probably more in the construction of prediction intervals for the mean of additional observations—in so far as the mean is of interest.

3. USE OF \( F \) RATIO FOR TESTING THAT \( E(\bar{d}) \) IS A ZERO VECTOR

Instead of changing the null hypothesis \( \beta_2 = \beta_1 = \beta \) to the hypothesis \( E(\bar{d}) = 0 \), consider the quadratic form \( d' (\text{Cov} \ d)^{-1} d \). It follows from (4) and (6) that

\[
d' (\text{Cov} \ d)^{-1} d = \beta_2 X_2' - \beta_1 X_2' [I + X_2(X_1'X_1)^{-1} X_2']^{-1} [X_2 \beta_2 - X_2 \beta_1] \frac{1}{\sigma^2} + [\varepsilon_1' \varepsilon_2'] \left[ -X_1(X_1'X_1)^{-1} X_2' [I + X_2(X_1'X_1)^{-1} X_2']^{-1} [-X_2(X_1'X_1)^{-1} X_1'] I \right] \frac{\varepsilon_1}{\varepsilon_2} \frac{1}{\sigma^2}
\]

The last term is a quadratic form in \( [\varepsilon_1' \varepsilon_2'] \) with rank \( m \)—note that \( [I + X_2 (X_1'X_1)^{-1} X_2']^{-1} \) is \( m \) by \( m \). It will be equal to \( d' (\text{Cov} \ d)^{-1} d \) under the null hypothesis that \( \beta_2 = \beta_1 \), as can easily be seen from (12). Therefore under the null hypothesis, \( d' (\text{Cov} \ d)^{-1} d \) will follow \( \chi^2(m) \) distribution; whereas under the alternative hypothesis \( \beta_2 \neq \beta_1 \), \( d' (\text{Cov} \ d)^{-1} d \) will follow a non-central \( \chi^2 \) distribution.

\(^2\) I am indebted to Robert Solow for pointing out this reference and the errors therein.
It is well known that the square of the standard error from the first regression, \( s_1^2 \), times \((n - \beta)/\sigma^2\), follows \( \chi^2(n - \beta) \). This \( \chi^2(n - \beta) \) is independent of \( d = y_2 - X_2b_1 \), \( s_1^2 \) is independent of \( b_1 \) and is certainly independent of \( y_2 \). Therefore, under the null hypothesis, the ratio

\[
\frac{d'(\text{Cov} \ d)^{-1} d}{s_1^2 (n - \beta)/\sigma^2} \cdot \frac{1}{s_1^2 m} = \frac{d'[I + X_2(X_2'X_2)^{-1} X_2']^{-1} d}{s_1^2 m}
\]

will follow \( F(m, n - \beta) \). Since the numerator of (13) will have a non-central \( \chi^2 \) distribution when \( \beta_2 \neq \beta_1 \), the upper-tail \( F \) test can appropriately be used. Clearly the test (13) reduces to the prediction interval (8) when \( m = 1 \).

4. RELATIONSHIPS OF PREDICTION INTERVAL AND ANALYSIS OF COVARIANCE TO THEORY OF LINEAR HYPOTHESES

This section shows the relationships among the \( F \) test of (13), the prediction interval for one additional observation, and the analysis of covariance (for \( m > \beta \)). All three methods are special applications of the theory of testing general linear hypotheses. It will therefore be convenient to summarize first the theory of linear hypotheses as applied to testing the homogeneity of the (entire) sets of coefficients in two regressions. The size \( m \) of the second sample will first be assumed to be larger than \( \beta \), and then reduced to one.

In our context, the model of general linear hypotheses takes the form

\[
y_1 = X_1\beta_1 + 0 \beta_2 + \epsilon_1
\]
\[
y_2 = 0 \beta_1 + X_2\beta_2 + \epsilon_2
\]

or

\[
\begin{bmatrix}
y_1 \\
y_2
\end{bmatrix} =
\begin{bmatrix}
X_1 & 0 \\
0 & X_2
\end{bmatrix}
\begin{bmatrix}
\beta_1 \\
\beta_2
\end{bmatrix} +
\begin{bmatrix}
\epsilon_1 \\
\epsilon_2
\end{bmatrix}.
\]

Under the null hypothesis \( (H_0: \beta_1 = \beta_2 = \beta) \), the model becomes

\[
\begin{bmatrix}
y_1 \\
y_2
\end{bmatrix} =
\begin{bmatrix}
X_1 \\
X_2
\end{bmatrix} \beta +
\begin{bmatrix}
\epsilon_1 \\
\epsilon_2
\end{bmatrix}.
\]

The sum of squares of the residuals under \( H_0 \) will be shown to equal the sum of squares of residuals under the alternative hypothesis \( (H_a: \beta_1 \neq \beta_2) \) plus the sum of squares of the deviations between the two sets of estimates of \( y \) under these two hypotheses. The ratio between the latter two sums, adjusted for their numbers of degrees of freedom, will be shown to follow an \( F \) distribution if the null hypothesis is true.

The developments here follow, and are special applications of, Kempthorne [6].
If the null hypothesis is true, the least-squares (also maximum likelihood) estimator of \( \beta \), denoted by \( b_o \), is

\[
(16) \quad b_o = \left( \frac{X_1 X_2}{X_2} \right)^{-1} \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} = [X_1 X_1 + X_2 X_2]^{-1} \frac{X_1 X_2}{X_2} \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} = \beta + [X_1 X_1 + X_2 X_2]^{-1} \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}.
\]

The residuals from this regression are:

\[
(17) \quad \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} - \frac{X_1}{X_2} b_o = \frac{X_1}{X_2} \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix} - \frac{X_1}{X_2} \beta
\]

\[
= \left[ I - \frac{X_1}{X_2} \right] \begin{bmatrix} X_1 X_1 + X_2 X_2 \end{bmatrix}^{-1} \begin{bmatrix} X_1 X_2 \\ X_2 \end{bmatrix} \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}.
\]

The sum of squares of the residuals under \( H_o \) can be written as

\[
(18) \quad \|\gamma_1 - \frac{X_1}{X_2} b_o\|^2 = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}^t \left[ I - \frac{X_1}{X_2} \right] \begin{bmatrix} X_1 X_1 + X_2 X_2 \end{bmatrix}^{-1} \begin{bmatrix} X_1 X_2 \\ X_2 \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}^t \left[ I - \frac{X_1}{X_2} \right] \begin{bmatrix} X_1 X_1 + X_2 X_2 \end{bmatrix}^{-1} \begin{bmatrix} X_1 X_2 \\ X_2 \end{bmatrix} \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}.
\]

Since these residuals are from a regression of \( n + m \) observations on \( \hat{p} \) explanatory variables, the quadratic form (18) in the \( \epsilon \)'s has rank \( n + m - \hat{p}. \)

If the alternative hypothesis \( (H_a: \beta_1 \neq \beta_2) \) is true, we are back to the model (14), and the least-squares estimators of \( \beta_1 \) and \( \beta_2 \) are

\[
(19) \quad \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} X_1 X_1 & 0 \\ 0 & X_2 X_2 \end{bmatrix}^{-1} \begin{bmatrix} X_1 X_1 & 0 \\ 0 & X_2 X_2 \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} = \begin{bmatrix} (X_1 X_1)^{-1} X_1 \gamma_1 \\ (X_2 X_2)^{-1} X_2 \gamma_2 \end{bmatrix}.
\]

The residuals under \( H_a \) will be

\[
(20) \quad \begin{bmatrix} y_1 - X_1 b_1 \\ y_2 - X_2 b_2 \end{bmatrix} = \begin{bmatrix} [I - X_1 (X_1 X_1)^{-1} X_1] \epsilon_1 \\ [I - X_2 (X_2 X_2)^{-1} X_2] \epsilon_2 \end{bmatrix}.
\]

Similarly, the sum of squares of these residuals will be

\[
(21) \quad \|y_1 - X_1 b_1\|^2 = \|y_2 - X_2 b_2\|^2
\]

\[
= \epsilon_1^t [I - X_1 (X_1 X_1)^{-1} X_1] \epsilon_1 + \epsilon_2^t [I - X_2 (X_2 X_2)^{-1} X_2] \epsilon_2.
\]

Since the last two quadratic forms have ranks \( n - \hat{p} \) and \( m - \hat{p} \) respectively, and since \( \epsilon_1 \) and \( \epsilon_2 \) are independent, the rank of the quadratic form (21) will be \( n + m - 2\hat{p} \).

\[\text{\footnote{For a proof of this, see Kempthorne\cite{[6]}.}}\]
Now the sum of squares (18) under $H_o$ will be decomposed into the sum of squares (21) under $H_o$ plus the sum of squares of the differences

\[ [X_1 b_1 - X_1 b_o] \text{ and } [X_2 b_2 - X_2 b_o]. \]

First start from the identity

\[
\begin{bmatrix}
y_1 - X_1 b_o \\
y_2 - X_2 b_o
\end{bmatrix} =
\begin{bmatrix}
y_1 - X_1 b_1 \\
y_2 - X_2 b_2
\end{bmatrix} +
\begin{bmatrix}
X_1 b_1 - X_1 b_o \\
X_2 b_2 - X_2 b_o
\end{bmatrix}.
\]

Summing the squares of the elements on both sides of (22) gives

\[
\begin{bmatrix}
y_1 - X_1 b_o \\
y_2 - X_2 b_o
\end{bmatrix}^2 =
\begin{bmatrix}
y_1 - X_1 b_1 \\
y_2 - X_2 b_2
\end{bmatrix}^2 +
\begin{bmatrix}
X_1 b_1 - X_1 b_o \\
X_2 b_2 - X_2 b_o
\end{bmatrix}^2
\]

because the cross-product term on the right side of (23) can easily be seen to be zero. To economize space, (23) will also be written as

\[
Q_1 = Q_2 + Q_3.
\]

We will proceed to show that the rank of the quadratic form $Q_3$ can at most be $p$. From (16) and (19), it follows that

\[
[X_1' X_1 + X_2' X_2] b_o = X_1' y_1 + X_2' y_2 = X_1' X_1 b_1 + X_2' X_2 b_2
\]

which implies

\[
b_2 - b_o = - (X_2' X_2)^{-1} X_1' X_1 (b_1 - b_o).
\]

Substituting (26) into $Q_3$, we have

\[
Q_3 = \| X_1 (b_1 - b_o) - X_2 (X_2' X_2)^{-1} X_1' X_1 (b_1 - b_o) \|^2
\]

\[
= [b_1 - b_o] [X_1 - X_1' X_1 (X_2' X_2)^{-1} X_2'] [X_1 - X_1' X_1 (X_2' X_2)^{-1} X_2'] [b_1 - b_o].
\]

(27) is a quadratic form in $b_1 - b_o$ and therefore cannot have rank higher than $p$. But $b_1 - b_o$ is a linear transformation of the $e_i$'s, as can be shown from (2) and (16):

\[
b_1 - b_o = \beta_1 - \beta + \{[(X_1' X_1)^{-1} X_1' 0] - [X_1' X_1 + X_2' X_2]^{-1} [X_1' X_2]\} [e_1].
\]

Under the null hypothesis $\beta_1 = \beta_2 = \beta$, $Q_3$ will thus be a quadratic form in the $e_i$'s with a maximum rank of $p$. From (28), we also see that $Q_3$ will tend to be larger when the null hypothesis is not true.

It has already been observed that the rank of $Q_2$ is $m + n - 2p$. Since the rank of $Q_1$ is smaller than or equal to the rank of $Q_2$ plus the rank of $Q_3$, the rank of $Q_3$ must be $p$. Under the null hypothesis $Q_2$ and $Q_3$ will be distributed independently as $\chi^2(m + n - 2p) \sigma^2$ and $\chi^2(p) \sigma^2$. While the
distribution of \( Q_3 \) is affected if \( H_o \) does not hold, \( Q_2 \) will have the same distribution regardless. We thus can test \( H_o \) by the \( F \) ratio

\[
F(\hat{\phi}, m + n - 2\hat{\phi}) = \frac{Q_3/\hat{\phi}}{Q_2/(m + n - 2\hat{\phi})}
= \frac{||X_1b_1 - X_1b_o||^2 + ||X_2b_2 - X_2b_o||^2}{||y_1 - X_1b_1||^2 + ||y_2 - X_2b_2||^2} \cdot \frac{(m + n - 2\hat{\phi})}{\hat{\phi}}.
\]

(29) is the standard analysis-of-covariance test when \( m > \hat{\phi} \).\(^5\)

A few remarks will suffice to indicate the application of the theory of linear hypotheses to the case \( m \leq \hat{\phi} \). Let us rewrite (23) as

\[
||y_1 - X_1b_1||^2 = ||y_1 - X_1b_1||^2 + ||X_1b_1 - X_1b_o||^2 + ||y_2 - X_2b_2||^2
+ ||X_2b_2 - X_2b_o||^2
\]

(30)

Our models for \( H_o \) and \( H_a \) are (15) and (14), as before. The sum of squares under \( H_o \) is clearly \( Q_1 \) whether \( m > \hat{\phi} \) or \( m \leq \hat{\phi} \). The sum of squares under \( H_a \) will become \( ||y_1 - X_1b_1||^2 \) when \( m \leq \hat{\phi} \)—this can be seen either by evaluating the sum of squares of the residuals from regression (14) or by noting that the residuals from the second sample will simply be zero. Regardless of the size \( m \), \( ||y_1 - X_1b_1||^2 \) will be distributed as \( \chi^2(n - \hat{\phi})\sigma^2 \) and will be independent of the sum of the other three terms on the right side of (30). The sum of these three terms equals

\[ ||X_1b_1 - X_1b_o||^2 + ||y_2 - X_2b_2||^2 \]

even if \( b_2 \) is undefined. When \( m \leq \hat{\phi} \), we can test \( H_o \) by the ratio

\[
F(m, n - \hat{\phi}) = \frac{||X_1b_1 - X_1b_o||^2 + ||y_2 - X_2b_2||^2}{||y_1 - X_1b_1||^2} \cdot \frac{(n - \hat{\phi})}{m}.
\]

(31)

When \( m > \hat{\phi} \), (31) remains valid. However, using (31) instead of (29) in this situation would amount to taking a part of \( Q_1 \), i.e., \( ||y_2 - X_2b_2||^2 \), which is not affected by the inequality between \( \beta_1 \) and \( \beta_2 \), and placing it in the numerator of the \( F \) ratio. This would reduce the power of the test.

The theory of linear hypotheses has now been applied to testing the homogeneity of two regressions. To provide a link between the analysis of covariance (29) and the prediction interval (8), we will point out that the test (13) in Section 3, including its special case (8), is identical with the test (31). The proof of this identity requires only the proof that

\[
d'[I + X_2(X_1'X_1)^{-1}X_2']^{-1}d = ||X_1b_1 - X_1b_o||^2 + ||y_2 - X_2b_o||^2.
\]

\(^5\) Additional references on the analysis of covariance include [1], [5], [9], and [10].
[1] is a special issue devoted to the analysis of covariance mainly for the design of experiments.
From (25), we deduce

\begin{equation}
(33) \quad b_1 = [I + (X_1X_1)^{-1} X_2 X_2] b_o - (X_1X_1)^{-1} X_2 y_2.
\end{equation}

Substitute (33) into \(d\):

\begin{equation}
(34) \quad d = y_2 - X_2 b_1 = [I + X_2 (X_1X_1)^{-1} X_2] [y_2 - X_2 b_o].
\end{equation}

Given (34), we evaluate the quadratic form

\begin{equation}
(35) \quad d' [I + X_2 (X_1X_1)^{-1} X_2]^{-1} d = [y_2' - b_o X_2'] [I + X_2 (X_1X_1)^{-1} X_2] [y_2 - X_2 b_o]
\end{equation}

\[= [y_2' - b_o X_2'] [y_2 - X_2 b_o] + [y_2 X_2 - b_o X_2 X_2] (X_1X_1)^{-1} [X_2 y_2 - X_2' X_2 b_o].\]

Our proof of (32) will be complete by observing the relationship, based on (25), that

\begin{equation}
(36) \quad X_2 y_2 - X_2' X_2 b_o = - [X_1X_1 b_1 - X_1X_1 b_o].
\end{equation}

5. Tests of Equality Between Subsets of Coefficients in Two Regressions

The results given so far, as summarized by (29) and (31), will now be extended to testing the equality between subsets of coefficients in two regressions. As before, we will first examine the case \(m > p\).

Under the alternative hypothesis, our model is

\begin{equation}
(37) \quad y_1 = X_1 \beta_1 + \varepsilon_1 = Z_1 \gamma_1 + W_1 \delta_1 + \varepsilon_1, \quad y_2 = X_2 \beta_2 + \varepsilon_2 = Z_2 \gamma_2 + W_2 \delta_2 + \varepsilon_2,
\end{equation}

or

\[
\begin{bmatrix}
\gamma_1 \\
\gamma_2
\end{bmatrix} =
\begin{bmatrix}
Z_1 & 0 & W_1 & 0 \\
0 & Z_2 & 0 & W_2
\end{bmatrix}
\begin{bmatrix}
\gamma_1 \\
\gamma_2 \\
\delta_1 \\
\delta_2
\end{bmatrix} +
\begin{bmatrix}
\varepsilon_1 \\
\varepsilon_2
\end{bmatrix},
\]

where the coefficients \(\beta_1\) are divided into \(\gamma_1\) and \(\delta_1\), the matrix \(X_1\) is correspondingly divided into \(Z_1\) and \(W_1\), and similarly for \(\beta_2\) and \(X_2\). Let \(\gamma_1\) and \(\gamma_2\) be column vectors of \(q\) elements each; \(\delta_1\) and \(\delta_2\) be column vectors of \(p - q\) elements each. The subsets of coefficients to be tested are \(\gamma_1\) and \(\gamma_2\).

The null hypothesis is \(\gamma_1 = \gamma_2 = \gamma\), implying the model

\begin{equation}
(38) \quad
\begin{bmatrix}
\gamma_1 \\
\gamma_2
\end{bmatrix} =
\begin{bmatrix}
Z_1 & 0 & W_1 & 0 \\
Z_2 & 0 & W_2
\end{bmatrix}
\begin{bmatrix}
\gamma \\
\delta_1 \\
\delta_2
\end{bmatrix} +
\begin{bmatrix}
\varepsilon_1 \\
\varepsilon_2
\end{bmatrix}.
\end{equation}
Under the null hypothesis, the least-squares estimators of the coefficients are

\[
\begin{bmatrix}
  c_0 \\
  d_{10} \\
  d_{20}
\end{bmatrix} = \begin{bmatrix}
  Z_1 Z_1 + Z_1 Z_2 & Z_1 W_1 & Z_2 W_2 \\
  W_1 Z_1 & W_1 W_1 & 0 \\
  W_2 Z_2 & 0 & W_2 W_2
\end{bmatrix}^{-1} \begin{bmatrix}
  Z_1 \\
  Z_2 \\
  W_1
\end{bmatrix} \begin{bmatrix}
  y_1 \\
  y_2
\end{bmatrix}.
\]

The sum of squares of the residuals under \( H_0 \), analogous to (18), will be

\[
\begin{bmatrix}
  \varepsilon_1 \\
  \varepsilon_2
\end{bmatrix} = \begin{bmatrix}
  Z_1 & W_1 & 0 \\
  Z_2 & 0 & W_2
\end{bmatrix} \begin{bmatrix}
  Z_1 Z_1 + Z_2 Z_2 & Z_1 W_1 & Z_2 W_2 \\
  W_1 Z_1 & W_1 W_1 & 0 \\
  W_2 Z_2 & 0 & W_2 W_2
\end{bmatrix}^{-1} \begin{bmatrix}
  Z_1 \\
  Z_2 \\
  W_1
\end{bmatrix} \begin{bmatrix}
  \varepsilon_1 \\
  \varepsilon_2
\end{bmatrix}
\]

with \( m + n - 2p + q \) degrees of freedom.

Under the alternative hypothesis \( \gamma_1 \neq \gamma_2 \), the least-squares estimators are

\[
\begin{bmatrix}
  c_1 \\
  c_2 \\
  d_1 \\
  d_2
\end{bmatrix} = \begin{bmatrix}
  Z_1 Z_1 & 0 & Z_1 W_1 & 0 \\
  0 & Z_2 Z_2 & 0 & Z_2 W_2 \\
  W_1 Z_1 & 0 & W_1 W_1 & 0 \\
  0 & W_2 Z_2 & 0 & W_2 W_2
\end{bmatrix}^{-1} \begin{bmatrix}
  Z_1 \\
  Z_2 \\
  W_1 \\
  W_2
\end{bmatrix} \begin{bmatrix}
  y_1 \\
  y_2
\end{bmatrix}.
\]

The sum of squares of the residuals under \( H_a \), which is identical with (21), will have \( m + n - 2p \) degrees of freedom.

As before, the sum of squares under \( H_0 \) can be broken up into the sum of squares under \( H_a \) plus the sum of squares of the differences between the two sets of estimates of \( \gamma \), namely,

\[
\begin{align*}
\begin{bmatrix}
  y_1 \\
  y_2
\end{bmatrix} - \begin{bmatrix}
  Z_1 & W_1 & 0 \\
  Z_2 & 0 & W_2
\end{bmatrix} \begin{bmatrix}
  c_0 \\
  d_{10} \\
  d_{20}
\end{bmatrix} & = \begin{bmatrix}
  y_1 \\
  y_2
\end{bmatrix} - \begin{bmatrix}
  Z_1 & W_1 & 0 \\
  0 & Z_2 & 0 \\
  0 & W_1 & 0
\end{bmatrix} \begin{bmatrix}
  c_1 \\
  c_2 \\
  d_1 \\
  d_2
\end{bmatrix} \\
+ \begin{bmatrix}
  y_1 \\
  y_2
\end{bmatrix} - \begin{bmatrix}
  Z_1 & W_1 & 0 \\
  0 & Z_2 & 0 \\
  0 & W_2 & 0
\end{bmatrix} \begin{bmatrix}
  c_0 \\
  d_{10} \\
  d_{20}
\end{bmatrix}
\end{align*}
\]

or

\[ Q^*_a = Q_2 + Q^*_3. \]

We will omit the proof that each of the cross-products on the right side of (42) is zero, but will indicate following identity which may be used in the proof:

\[
\begin{bmatrix}
  Z_1 & W_1 & 0 \\
  Z_2 & 0 & W_2
\end{bmatrix} = \begin{bmatrix}
  Z_1 & 0 & W_1 \\
  0 & Z_2 & 0 \\
  0 & W_1 & W_2
\end{bmatrix} \begin{bmatrix}
  I & 0 & 0 \\
  I & 0 & 0 \\
  0 & I & 0
\end{bmatrix}.
\]
Given that the ranks of $Q^*$ and $Q_2$ are respectively $m + n - 2p + g$ and $m + n - 2p$, the rank of $Q^*_g$ must be $g$ once it can be shown that it is at most $q$. To show the maximum rank of $Q^*_g$, we first define $y_1$, as the residuals of the regression of $y_1$ on $W_1$, $Z_1$, as the residuals of $Z_1$ on $W_1$, and similarly for $y_2$, and $Z_2$. Then it can be proved that

$$
\begin{bmatrix}
  c_1 \\
  c_2
\end{bmatrix} = \begin{bmatrix}
  (Z_1' Z_1)^{-1} Z_1' y_1 \\
  (Z_2' Z_2)^{-1} Z_2' y_2
\end{bmatrix}
$$

and

$$
\begin{bmatrix}
  d_1 \\
  d_2
\end{bmatrix} = \begin{bmatrix}
  (W_1' W_1)^{-1} W_1' y_1 \\
  (W_2' W_2)^{-1} W_2' y_2
\end{bmatrix} - \begin{bmatrix}
  (W_1' W_1)^{-1} W_1' Z_1 c_1 \\
  (W_2' W_2)^{-1} W_2' Z_2 c_2
\end{bmatrix}.
$$

Rather than digressing to complete the proofs of (44) and (45), we simply indicate that essentially they involve partitioning the matrix of the cross-products of the explanatory variables in (41) into four blocks and then inverting the partitioned matrix. The same method will also prove

$$
c_0 = [Z_1', Z_1, + Z_2', Z_2, ]^{-1} [Z_1', Z_2, ] [y_1',]
$$

and

$$
\begin{bmatrix}
  d_{10} \\
  d_{20}
\end{bmatrix} = \begin{bmatrix}
  (W_1' W_1)^{-1} W_1' y_1 \\
  (W_2' W_2)^{-1} W_2' y_2
\end{bmatrix} - \begin{bmatrix}
  (W_1' W_1)^{-1} W_1' Z_1 c_0 \\
  (W_2' W_2)^{-1} W_2' Z_2 c_0
\end{bmatrix}.
$$

Using (45) and (47), we can rewrite the vector of the differences between the estimates of $y$ under $H_a$ and under $H_0$:

$$
\begin{bmatrix}
  Z_1 & 0 & W_1 & 0 \\
  0 & Z_2 & 0 & W_2
\end{bmatrix}
\begin{bmatrix}
  c_1 \\
  c_2
\end{bmatrix} - \begin{bmatrix}
  Z_1 & 0 & W_1 & 0 \\
  Z_2 & 0 & W_2
\end{bmatrix}
\begin{bmatrix}
  d_{10} \\
  d_{20}
\end{bmatrix}

= \begin{bmatrix}
  I - (W_1 W_1 W_1)^{-1} W_1 \\
  0 & W_2 W_2 W_2)^{-1} W_2
\end{bmatrix}
\begin{bmatrix}
  Z_1(c_1 - c_0), \\
  Z_2(c_2 - c_0)
\end{bmatrix}.
$$

The developments from now on will correspond to the developments of (25), (26), (27), and (28) in Section 4. From (44) and (46), it follows that

$$
[Z_1, Z_1, + Z_2, Z_2, ] c_0 = Z_1, y_1, + Z_2, y_2, = Z_1, Z_1, c_1 + Z_2, Z_2, c_2
$$

which corresponds to (25). (49) can be used to express $c_2 - c_0$ as a linear transformation of $c_1 - c_0$, a transformation similar to the one in (26). Replacing $c_2 - c_0$ in (48) by this transformation of $c_1 - c_0$, we will observe that $Q^*$ is a quadratic form in $c_1 - c_0$, with maximum rank $q$. A further step analogous to (28) will show that $Q^*$ tends to be larger when $\gamma_1 \neq \gamma_2$. We therefore have
\begin{equation}
F(q, m + n - 2\phi) = \frac{Q^*_q}{Q_2/(m + n - 2\phi)}
= \frac{||Z_1c_1 + W_1d_1 - Z_1c_0 - W_1d_{10}||^2 + ||Z_2c_2 + W_2d_2 - Z_2c_0 - W_2d_{20}||^2}{||y_1 - Z_1c_1 - W_1d_1||^2 + ||y_2 - Z_2c_2 - W_2d_2||^2} \times \frac{(m + n - 2\phi)}{q}.
\end{equation}

The remaining case for examination is \((\phi - q) \leq m \leq \phi\). As long as \(m \geq (\phi - q)\), least-squares estimators under \(H_0\) can be obtained by (39), but not when \(m < (\phi - q)\) because \(W_1^2W_2\) will then be singular. The sum of squares of the residuals under \(H_0\) will still be \(Q^*_1\). The sum of squares of the residuals under \(H_a\) will be \(||y_1 - Z_1c_1 - W_1d_1||^2\) with \(n - \phi\) degrees of freedom. Corresponding to (31), we have
\begin{equation}
F(m - \phi + q, n - \phi) = \frac{||Z_1c_1 + W_1d_1 - Z_1c_0 - W_1d_{10}||^2 + ||Z_2c_2 + W_2d_2 - Z_2c_0 - W_2d_{20}||^2}{||y_1 - Z_1c_1 - W_1d_1||^2} \cdot \frac{(n - \phi)}{(m - \phi + q)}.
\end{equation}

The results of this paper, which are really contained in (50) and (51), can be summarized briefly. To test the equality between sets of coefficients in two linear regressions, we obtain the sum of squares of the residuals assuming the equality, and the sum of squares without assuming the equality. The ratio of the difference between these two sums to the latter sum, adjusted for the corresponding degrees of freedom, will be distributed as the \(F\) ratio under the null hypothesis. This latter sum of squares will be computed only from the first sample of \(n\) observations when the second sample is not large enough for computing a separate regression. We have attempted to show how the theory of general linear hypotheses is applied to our problem and how the prediction interval and the analysis of covariance are related to each other and to the theory of general linear hypotheses. While we have dealt with the comparison of coefficients in only two regressions, the proofs of (29) and (50) can obviously be generalized to the case of many regressions.

6. EXAMPLES

To illustrate how some of the tests given above are applied, one numerical example for each of (29) and (31) will now be provided. These examples originated in a study of the demand for automobiles in the United States [2]. We will not here go into the economic justifications of them as they are contained in the reference cited. The study utilizes annual observations on the following variables:
$X_t$, ownership of automobiles measured in "new-car equivalents" per capita at the end of year $t$. The unit is per cent of a new-car equivalent per capita.

$X_t$, purchase of new cars during year $t$, with the same unit of measurement as above.

$P_t$, relative price index of automobile stock, with 1937 as 100.

$I_{dt}$, real disposable income per capita in 1937 dollars.


A statistical demand function for automobile ownership computed from observations of 33 years from 1921 to 1953 is

$$X_t = -0.7247 - 0.048802 P_t + 0.025487 I_{et}, \quad R^2 = 0.895,$$

$$(0.004201) \quad (0.001747) \quad s = 0.618.$$

A statistical demand function for new purchase computed from observations of 28 years, from 1921 to 1953 but excluding 1942 to 1946, is

$$X^1_t = 0.07791 - 0.020127 P_t + 0.011699 I_{dt} - 0.23104 X_t - 1, \quad R^2 = 0.858,$$

$$(0.002648) \quad (0.001070) \quad (0.04719) \quad s = 0.308.$$

Four years after the study had been made, four additional observations were available for testing whether these demand functions remained stable over time. Since four observations are sufficient for computing a separate regression of the form (X1e), test (29) was used. To determine the stability of (4s), test (31) was used. The follow-up study is described more thoroughly in [3]. Before presenting the analysis of covariance (29) for the demand function (X1e), we exhibit here the estimated values of the dependent variable together with the deviations of the observed values from the estimated values.

<table>
<thead>
<tr>
<th>Year</th>
<th>Estimated</th>
<th>1954</th>
<th>1955</th>
<th>1956</th>
<th>1957</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_t$ estimated from (X1e)</td>
<td>12.665</td>
<td>12.993</td>
<td>13.328</td>
<td>13.025</td>
<td></td>
</tr>
<tr>
<td>$X_t$ observed minus estimated</td>
<td>-0.613</td>
<td>0.079</td>
<td>0.102</td>
<td>0.437</td>
<td></td>
</tr>
</tbody>
</table>

The residuals of the observed values from the estimated values are very small, as compared with the standard error of .618. They do not indicate any shifts in the pattern of demand for automobile ownership during the four years 1954 to 1957.

We will now proceed with the analysis of covariance (29). The method involved can be described very simply. Suppose that $n$ observations are used to estimate a regression with $p$ parameters ($p = 1$ coefficients plus one intercept). Suppose also that there are $m$ additional observations, and we are interested in deciding whether they are generated by the same regression model as the first $n$ observations. To perform the analysis of covariance, we need the following sums of squares:
\( A \), sum of squares of \( n + m \) deviations of the dependent variable from the regression estimated by \( n + m \) observations, with \( n + m - \phi \) degrees of freedom.

\( B \), sum of squares of \( n \) deviations of the dependent variable from the regression estimated by the first \( n \) observations, with \( n - \phi \) degrees of freedom.

\( C \), sum of squares of \( m \) deviations of the dependent variable from the regression estimated by the second \( m \) observations, with \( m - \phi \) degrees of freedom.

From (29), the ratio of \( (A - B - C)/\phi \) to \( (B + C)/(n + m - 2\phi) \) will be distributed as \( F(\phi, n + m - 2\phi) \) under the null hypothesis that both groups of observations belong to the same regression model. For testing the demand function (X1e), the sum of squares \( A \) is 10.1155, and \( B + C \) is 9.6130. The ratio \( F(3,26) \) is therefore 0.45. In order to interpret the new observations as coming from a different structure at the 5 per cent level of significance, \( F \) would have to be at least 2.98. Our impression from examining the four deviations that there was no change in structure is strongly confirmed.

The following is a comparison of the estimated and the observed values of the dependent variable of demand function (4s).

\[
\begin{array}{cccc}
\text{Year} & 1954 & 1955 & 1956 & 1957 \\
\hline
X^1_i \text{ estimated from (4s)} & 3.452 & 3.730 & 3.630 & 3.270 \\
X^1_i \text{ observed minus estimated} & -.044 & .608 & -.087 & .226 \\
\end{array}
\]

Again, inspection of the residuals reveals that they are not large relative to 0.308, the standard error of estimate for (4s). The year 1955 is an exception, where we find the residual to be twice as large as the standard error. To apply test (31), we compute the sum of squares \( A \) of the 32 deviations from the regression including the four new observations, with \( 32 - 4 \) or 28 degrees of freedom. \( A \) turns out to be 2.6444 numerically. The sum of squares \( B \) of the 28 deviations from the regression of the original set of observations turns out to be 2.2818, with 24 degrees of freedom. The sum of squares \( C \) vanishes as long as the number \( m \) of new observations does not exceed the number of parameters \( \phi \). According to (31), the \( F \) ratio is the ratio of \( (A - B)/4 \) to \( B/24 \), or 0.95 numerically. Therefore, we accept the null hypothesis that automobile purchases in the years 1954 to 1957 were governed by the same relationship as before.

Cornell University
REFERENCES