CHAPTER 9

Multiple Regression: Special Topics

The topics in this chapter are all useful on certain occasions, but most or all of them can probably be passed by on a first reading. Thus it is convenient to group them together at a point in the book where the necessary prerequisites have been established.

9.1. TESTING A GENERAL LINEAR HYPOTHESIS

Experimenters sometimes postulate models that are more general than they hope they need. For example, suppose an experimenter is involved with a response Y and two predictors X_1 and X_2 and has a set of data (Y_i, X_{1i}, X_{2i}), $i = 1, 2, \ldots, n$. She suspects that, although X_1 and X_2 both affect Y, the single predictor of importance is really the difference $X_1 - X_2$. If both X's are needed, she will want to fit the model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon, \tag{9.1.1}$$

but, if her suspicion is correct, the model

$$Y = \beta_0 + \beta(X_1 - X_2) + \epsilon \tag{9.1.2}$$

would be good enough. How can she check? Essentially, she has asked the question: "Could it be, in Eq. (9.1.1), that $\beta_1 = -\beta_2$ ($= \beta$, say)?" Or, alternatively, "Is $\beta_1 + \beta_2 = 0$?" She thus will want to test the null hypothesis $H_0 : \beta_1 + \beta_2 = 0$ versus the alternative $H_1 : \beta_1 + \beta_2 \neq 0$. Because H_0 involves a statement about a linear combination of the β's, we call it a *linear hypothesis*.

Linear hypotheses typically arise *from the knowledge of the experimenter and his/her conjectures about possible models*. They can also arise from a consulting statistician, if he/she is deeply enough involved in the project to understand it at such a level. Ideally, the statistician *should* be that deeply involved, but in practice this does not always happen.

A linear hypothesis can also consist of more than one statement about the β's. We now provide some additional examples of linear hypotheses, explain generally how one is tested, and illustrate the procedure with a simple numerical example, H_1 is always the statement that H_0 is not true in some way, and so is not specifically mentioned in the examples.

(We note that the "extra sum of squares" principle of Section 6.1 is a special case of the work in this section.)

Example 1. Model: $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$.

$$H_0: \beta_1 = 0,$$
$$\beta_2 = 0 \qquad \text{(two linear functions, independent).}$$

(By "independent" we mean *linearly* independent, so that one statement cannot be obtained as a linear combination of other statements in the group.)

Example 2. Model: $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$.

$$\begin{aligned} H_0: \beta_1 &= 0, \\ \beta_2 &= 0, \\ &\vdots \\ \beta_k &= 0 \qquad (k \text{ linear functions, all independent}). \end{aligned}$$

Example 3. Model: $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$.

$$\begin{aligned} H_0: \beta_1 - \beta_2 &= 0, \\ \beta_2 - \beta_3 &= 0, \\ &\vdots \\ \beta_{k-1} - \beta_k &= 0 \qquad (k - 1 \text{ linear functions, independent}). \end{aligned}$$

Note that this expresses the hypothesis

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_k = \beta, \text{ say.}$$

Example 4 (*General Case*). Model: $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$.

$$\begin{aligned} H_0: c_{10}\beta_0 + c_{11}\beta_1 + c_{12}\beta_2 + \cdots + c_{1k}\beta_k &= 0, \\ c_{20}\beta_0 + c_{21}\beta_1 + c_{22}\beta_2 + \cdots + c_{2k}\beta_k &= 0, \\ &\vdots \\ c_{m0}\beta_0 + c_{m1}\beta_1 + c_{m2}\beta_2 + \cdots + c_{mk}\beta_k &= 0. \end{aligned}$$

In this hypothesis there are m linear functions of $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$, all of which may not be independent. H_0 can be expressed in matrix form as

$$H_0: \mathbf{C}\boldsymbol{\beta} = \mathbf{0},$$

where

$$\mathbf{C} = \begin{bmatrix} c_{10} & c_{11} & c_{12} & \cdots & c_{1k} \\ c_{20} & c_{21} & c_{22} & \cdots & c_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ c_{m0} & c_{m1} & c_{m2} & \cdots & c_{mk} \end{bmatrix}, \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}.$$

We shall suppose in what follows that the m functions are *dependent* and that the last $(m - q)$ of them depend on the first q; that is, if we had these first q independent functions, we could take linear combinations of them to form the other $(m - q)$ linear functions.

We have seen earlier how it is possible to test hypotheses of the forms in Examples 1 and 2. We now explain how more general hypotheses can be tested.

Testing a General Linear Hypothesis $\mathbf{C}\boldsymbol{\beta} = 0$

Suppose that the model under consideration, assumed correct, is

$$E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta},$$

where $\mathbf{Y}$ is $(n \times 1)$, $\mathbf{X}$ is $(n \times p)$, and $\boldsymbol{\beta}$ is $(p \times 1)$. If $\mathbf{X}'\mathbf{X}$ is nonsingular we can estimate $\boldsymbol{\beta}$ as

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

The residual sum of squares for this analysis is given, as we have seen, by

$$\text{SSE} = \mathbf{Y}'\mathbf{Y} - \mathbf{b}'\mathbf{X}'\mathbf{Y}.$$

This sum of squares has $(n - p)$ degrees of freedom. The linear hypothesis to be tested, $H_0: \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$, provides q independent conditions on the parameters $\beta_0, \beta_1, \ldots, \beta_k$, on the assumptions (mentioned above) that $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ represents m equations, of which only q are independent. We can use the q independent equations to solve for q of the β's in terms of the other $p - q$ of them. Substituting these solutions back into the original model provides us with a reduced model of, say,

$$E(\mathbf{Y}) = \mathbf{Z}\boldsymbol{\alpha},$$

where $\boldsymbol{\alpha}$ is a vector of parameters to be estimated. There will be $p - q$ of these parameters. The right-hand side $\mathbf{Z}\boldsymbol{\alpha}$, where $\mathbf{Z}$ is $n \times (p - q)$ and $\boldsymbol{\alpha}$ is $(p - q) \times 1$, represents the result of substituting into $\mathbf{X}\boldsymbol{\beta}$ for the dependent β's.

We can now estimate the parameter vector $\boldsymbol{\alpha}$ in the new model by

$$\mathbf{a} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y},$$

if $\mathbf{Z}'\mathbf{Z}$ is nonsingular, and can obtain a new residual sum of squares for this regression of

$$\text{SSW} = \mathbf{Y}'\mathbf{Y} - \mathbf{a}'\mathbf{Z}'\mathbf{Y}.$$

This sum of squares has $(n - p + q)$ degrees of freedom.

Since fewer parameters are involved in this second analysis, SSW will always be larger than SSE. The difference SSW $-$ SSE is called the *sum of squares due to the hypothesis* $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ and has $(n - p + q) - (n - p) = q$ degrees of freedom. A test of the hypothesis $H_0: \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ is now made by considering the ratio

$$\left(\frac{\text{SSW} - \text{SSE}}{q}\right) \Bigg/ \left(\frac{\text{SSE}}{n - p}\right)$$

and referring it to the $F(q, n - p)$ distribution in the usual manner. If the errors are normally distributed and independent, this is an exact test.

The appropriate test for Examples 1 and 2 [already given as Eq. (5.3.2) where $q = k = p - 1$] is a special case of this. The reduced model in both cases consists of

$$E(\mathbf{Y}) = \mathbf{1}\beta_0$$

where $\mathbf{1}' = (1, 1, \ldots, 1)$ is a vector of all ones. Another way of writing this model is

$$E(Y_i) = \beta_0, \qquad i = 1, 2, \ldots, n.$$

Since $b_0 = \overline{Y}$, SSW $= \mathbf{Y}'\mathbf{Y} - n\overline{Y}^2$ with $(n - 1)$ degrees of freedom, whereas SSE $=$

$\mathbf{Y'Y} - \mathbf{b'X'Y}$ with $(n - k - 1)$ degrees of freedom. So the ratio for the test $\beta_1 = \beta_2 = \cdots = \beta_k = 0$ (for Example 2; when $k = 2$, we have Example 1) is simply

$$\left(\frac{\mathbf{b'X'Y} - nY^2}{k}\right) \Big/ \left(\frac{\mathbf{Y'Y} - \mathbf{b'X'Y}}{n-k-1}\right)$$

and this is referred to the $F(k, n - k - 1)$ distribution. This is exactly the procedure of Eq. (5.3.2) with $k = p - 1$, $\nu = n - k - 1$, and $s^2 = \mathrm{MS}_E = \mathrm{SSE}/\nu$.

We shall now illustrate the use of the procedure in a simple but not so typical case.

Worked Example. Given the model $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$, test the hypothesis $H_0: \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$, where

$$\mathbf{Y}' = (1, 4, 8, 9, 3, 8, 9),$$

$$\boldsymbol{\beta}' = (\beta_0, \beta_1, \beta_2, \beta_{11}),$$

$$\mathbf{X} = \begin{array}{c} \begin{array}{cccc} 1 & X & X_2 & X_1^2 \end{array} \\ \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 2 & 0 \end{bmatrix} \end{array},$$

and

$$\mathbf{C} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & -1 & 1 \\ 0 & 2 & -2 & 3 \end{bmatrix}.$$

Solution. We first find the residual sum of squares SSE when the original model, of form $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2$, is fitted. We find

$$(\mathbf{X'X})^{-1} = \begin{bmatrix} 7 & 0 & 3 & 4 \\ 0 & 4 & 0 & 0 \\ 3 & 0 & 9 & 0 \\ 4 & 0 & 0 & 4 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & -\frac{1}{6} & -\frac{1}{2} \\ 0 & \frac{1}{4} & 0 & 0 \\ -\frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} \\ -\frac{1}{2} & 0 & \frac{1}{6} & \frac{3}{4} \end{bmatrix},$$

$$\mathbf{X'Y} = \begin{bmatrix} 42 \\ 4 \\ 38 \\ 22 \end{bmatrix}, \qquad \mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'Y} = \begin{bmatrix} \frac{11}{3} \\ 1 \\ 3 \\ \frac{11}{6} \end{bmatrix}, \qquad \begin{array}{l} \mathbf{b'X'Y} = 312.33 \\ \mathbf{Y'Y} = 316 \end{array}$$

$$\mathrm{SSE} = 316 - 312.33 = 3.67.$$

The equations for the null hypothesis $H_0\colon \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ are

$$\begin{aligned} \beta_{11} &= 0, \\ \beta_1 - \beta_2 &= 0, \\ \beta_1 - \beta_2 + \beta_{11} &= 0, \\ 2\beta_1 - 2\beta_2 + 3\beta_{11} &= 0. \end{aligned}$$

The hypothesis can be more simply expressed as $H_0\colon \beta_{11} = 0$, $\beta_1 = \beta_2 = \beta$, say, since the third and fourth equations are linear combinations of the first and second equations.

Substituting these conditions in the model gives a reduced model

$$E(Y) = \beta_0 + \beta(X_1 + X_2) = \alpha_0 + \alpha Z,$$

where

$$\alpha_0 = \beta_0, \qquad \alpha = \beta, \qquad Z = X_1 + X_2.$$

Thus

$$\mathbf{Z} = \begin{bmatrix} 1 & (-1-1) \\ 1 & (1-1) \\ 1 & (-1+1) \\ 1 & (1+1) \\ 1 & (0+0) \\ 1 & (0+1) \\ 1 & (0+2) \end{bmatrix} = \begin{bmatrix} 1 & -2 \\ 1 & 0 \\ 1 & 0 \\ 1 & 2 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$$

$$\mathbf{Z}'\mathbf{Y} = \begin{bmatrix} 42 \\ 42 \end{bmatrix}, \qquad (\mathbf{Z}'\mathbf{Z})^{-1} = \begin{bmatrix} 7 & 3 \\ 3 & 13 \end{bmatrix}^{-1} = \frac{1}{82}\begin{bmatrix} 13 & -3 \\ -3 & 7 \end{bmatrix},$$

$$\mathbf{a} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y} = \frac{21}{41}\begin{bmatrix} 10 \\ 4 \end{bmatrix}, \qquad \mathbf{a}'\mathbf{Z}'\mathbf{Y} = 301.17,$$

$$\text{SSW} = 316 - 301.17 = 14.83.$$

Now $p = 4$, $n = 7$, $q = 2$, $n - p = 3$, and

$$\text{SSW} - \text{SSE} = 14.83 - 3.67 = 11.16 = \text{SS due to the hypothesis.}$$

The appropriate test statistic for H_0 is thus $(11.16/2) \div 3.67/3 = 4.56$. Since $F(2, 3, 0.95) = 9.55$, we *do not* reject H_0. Since the original model was $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2$ and the hypothesis *not* rejected implies $\beta_{11} = 0$, $\beta_1 = \beta_2 = \beta$, a more plausible model would be $E(Y) = \beta_0 + \beta(X_1 + X_2)$.

9.2. GENERALIZED LEAST SQUARES AND WEIGHTED LEAST SQUARES

It sometimes happens that some of the observations used in a regression analysis are "less reliable" than others. What this usually means is that the variances of the observations are not all equal; in other words the nonsingular matrix $\mathbf{V}(\boldsymbol{\epsilon})$ is not of the form $\mathbf{I}\sigma^2$ but is diagonal with unequal diagonal elements. It may also happen, in some problems, that the off-diagonal elements of $\mathbf{V}(\boldsymbol{\epsilon})$ are not zero, that is, the observations are correlated.

When either or both of these events occur, the ordinary least squares estimation formula $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ does not apply and it is necessary to amend the procedures for obtaining estimates. The basic idea is to transform the observations $\mathbf{Y}$ to other variables $\mathbf{Z}$, which *do* appear to satisfy the usual tentative assumptions [that $\mathbf{Z} = \mathbf{Q}\boldsymbol{\beta} + \mathbf{f}$, $E(\mathbf{f}) = \mathbf{0}$, $V(\mathbf{f}) = \mathbf{I}\sigma^2$, and, for F-tests and confidence intervals to be valid, that $\mathbf{f} \sim N(\mathbf{0}, \mathbf{I}\sigma^2)$] and to then apply the usual analysis to the variables so obtained. The estimates can then be reexpressed in terms of the original variables $\mathbf{Y}$. We shall describe how the usual regression procedures are changed. Suppose the model under consideration is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{9.2.1}$$

where

$$\mathbf{E}(\boldsymbol{\epsilon}) = \mathbf{0}, \quad \mathbf{V}(\boldsymbol{\epsilon}) = \mathbf{V}\sigma^2, \quad \text{and} \quad \boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{V}\sigma^2). \tag{9.2.2}$$

It can be shown that it is possible to find a nonsingular symmetric matrix $\mathbf{P}$ such that

$$\mathbf{P}'\mathbf{P} = \mathbf{P}\mathbf{P} = \mathbf{P}^2 = \mathbf{V}. \tag{9.2.3}$$

Let us write

$$\mathbf{f} = \mathbf{P}^{-1}\boldsymbol{\epsilon}, \qquad \text{so that } E(\mathbf{f}) = \mathbf{0}. \tag{9.2.4}$$

Now it is a fact that, if $\mathbf{f}$ is a vector random variable such that $E(\mathbf{f}) = 0$, then $E(\mathbf{f}\mathbf{f}') = \mathbf{V}(\mathbf{f})$, where the expectation is taken separately for every term in the square $n \times n$ matrix $\mathbf{f}\mathbf{f}'$. Thus

$$\begin{aligned} \mathbf{V}(\mathbf{f}) = E(\mathbf{f}\mathbf{f}') &= E(\mathbf{P}^{-1}\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\mathbf{P}^{-1}), \qquad \text{since } (\mathbf{P}^{-1})' = \mathbf{P}^{-1} \\ &= \mathbf{P}^{-1}E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}')\mathbf{P}^{-1} \\ &= \mathbf{P}^{-1}\mathbf{P}\mathbf{P}\mathbf{P}^{-1}\sigma^2 \\ &= \mathbf{I}\sigma^2. \end{aligned} \tag{9.2.5}$$

It is also true that $\mathbf{f} \sim N(\mathbf{0}, \mathbf{I}\sigma^2)$; that is, $\mathbf{f}$ is normally distributed, since the elements of $\mathbf{f}$ consist of linear combinations of the elements of $\boldsymbol{\epsilon}$, which is itself normally distributed.

Thus if we premultiply Eq. (9.2.1) by $\mathbf{P}^{-1}$ we obtain a new model

$$\mathbf{P}^{-1}\mathbf{Y} = \mathbf{P}^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{P}^{-1}\boldsymbol{\epsilon} \tag{9.2.6}$$

or

$$\mathbf{Z} = \mathbf{Q}\boldsymbol{\beta} + \mathbf{f} \tag{9.2.7}$$

with an obvious notation. It is now clear that we can apply the basic least squares theory to Eq. (9.2.7) since $E(\mathbf{f}) = \mathbf{0}$ and $\mathbf{V}(\mathbf{f}) = \mathbf{I}\sigma^2$. The residual sum of squares is

$$\mathbf{f}'\mathbf{f} = \boldsymbol{\epsilon}'\mathbf{V}^{-1}\boldsymbol{\epsilon} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}). \tag{9.2.8}$$

The normal equations $\mathbf{Q}'\mathbf{Q}\mathbf{b} = \mathbf{Q}'\mathbf{Z}$ become

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} \tag{9.2.9}$$

with solution

$$\mathbf{b} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} \tag{9.2.10}$$

when the matrix just inverted is nonsingular. The regression sum of squares is

$$\mathbf{b}'\mathbf{Q}'\mathbf{Z} = \mathbf{Y}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} \tag{9.2.11}$$

and the total sum of squares is

$$\mathbf{Z}'\mathbf{Z} = \mathbf{Y}'\mathbf{V}^{-1}\mathbf{Y}. \tag{9.2.12}$$

The difference between Eqs. (9.2.12) and (9.2.11) provides the residual sum of squares. The sum of squares due to the mean is $(\Sigma Z_i)^2/n$, where Z_i are the n elements of the vector $\mathbf{Z}$. Note that, if we subtract this from Eq. (9.2.11), the remainder is not an extra sum of squares in the usual sense, because the transformed model no longer contains a β_0. Thus an appropriate base sum of squares to subtract here is one due to the first component of Eq. (9.2.7). The variance–covariance matrix of $\mathbf{b}$ is

$$\mathbf{V}(\mathbf{b}) = (\mathbf{Q}'\mathbf{Q})^{-1}\sigma^2 = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\sigma^2. \tag{9.2.13}$$

A joint confidence region for all the parameters can be obtained from

$$(\mathbf{b} - \boldsymbol{\beta})'\mathbf{Q}'\mathbf{Q}(\mathbf{b} - \boldsymbol{\beta}) = \left[\frac{p}{(n-p)}\right](\mathbf{Z}'\mathbf{Z} - \mathbf{b}'\mathbf{Q}'\mathbf{Z})F(p, n-p, 1-\alpha) \tag{9.2.14}$$

after substituting from Eqs. (9.2.11) and (9.2.12) and setting $\mathbf{Q} = \mathbf{P}^{-1}\mathbf{X}$, if so desired.

Generalized Least Squares Residuals

The residuals that must be checked are the estimates of $\mathbf{f} = \mathbf{P}^{-1}\boldsymbol{\epsilon}$. These residuals are given by

$$\mathbf{P}^{-1}(\mathbf{Y} - \hat{\mathbf{Y}}),$$

where $\hat{\mathbf{Y}} = \mathbf{X}\mathbf{b}$ and $\mathbf{b}$ is taken from Eq. (9.2.11). Thus these residuals are

$$\mathbf{P}^{-1}\{\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\}\mathbf{Y}. \tag{9.2.15}$$

A similar formula applies when $\mathbf{V}$ is estimated.

General Comments

We speak of *generalized least squares* when $\mathbf{V}$ is not a diagonal matrix, and of *weighted least squares* when it is. In the latter case, the observations are independent but have different variances so that

$$\mathbf{V}\sigma^2 = \begin{bmatrix} \sigma_1^2 & & & \mathbf{0} \\ & \sigma_2^2 & & \\ & & \ddots & \\ \mathbf{0} & & & \sigma_n^2 \end{bmatrix}$$

where some of the σ_i^2 may be equal.

In practical problems it is often difficult to obtain specific information on the form of $\mathbf{V}$ at first. For this reason it is sometimes necessary to make the (known to be erroneous) assumption $\mathbf{V} = \mathbf{I}$ and then attempt to discover something about the form of $\mathbf{V}$ by examining the residuals from the regression analysis.

If a generalized least squares analysis were called for but an ordinary least squares analysis were performed, the estimates obtained would still be unbiased but would not have minimum variance, since the minimum variance estimates are obtained from the correct generalized least squares analysis.

If standard least squares is used, then the estimates are obtained from $\mathbf{b}_0 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ and

$$E(\mathbf{b}_0) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}$$

but

$$\begin{aligned}\mathbf{V}(\mathbf{b}_0) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[\mathbf{V}(\mathbf{Y})]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\sigma^2.\end{aligned}$$

We recall from Eq. (9.2.13) that if the correct analysis is performed,

$$\mathbf{V}(\mathbf{b}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\sigma^2$$

and, in general, elements of this matrix would provide smaller variances both for individual coefficients and for linear functions of the coefficients.

Application to Serially Correlated Data

The major difficulty in applying generalized least squares methods is in finding $\mathbf{V}$ in Eq. (9.2.2). Suppose we wish to allow for serial correlation, for example. If the observations are listed in time order, the element V_{ij} of $\mathbf{V}$ would be ρ_l, where $l = |i - j|$, with $\rho_0 = 1$. To estimate ρ_l we could lag the observations by l steps and evaluate a correlation coefficient using Eq. (1.6.5), ignoring the unmatched overlap observations. These estimates are substituted into $\mathbf{V}$ to produce a $\hat{\mathbf{V}}$, which is used in formulas such as Eqs. (9.2.10) and (9.2.11). To analyze the residuals from this weighted fit we need the estimates of $\mathbf{f} = \mathbf{P}^{-1}\boldsymbol{\epsilon}$; see Eqs. (9.2.3) and (9.2.4). These estimates are thus

$$\hat{\mathbf{f}} = \hat{\mathbf{P}}^{-1}(\mathbf{Y} - \hat{\mathbf{Y}}), \tag{9.2.16}$$

where

$$\hat{\mathbf{P}}'\hat{\mathbf{P}} = \hat{\mathbf{V}} \tag{9.2.17}$$

and where $\hat{\mathbf{Y}}$ is fitted by generalized least squares, so that

$$\hat{\mathbf{Y}} = \mathbf{X}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{Y}. \tag{9.2.18}$$

In other words, the elements of

$$\hat{\mathbf{f}} = \hat{\mathbf{P}}^{-1}\{\mathbf{I} - \mathbf{X}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\}\mathbf{Y} \tag{9.2.19}$$

are examined.

(*Note:* This is essentially the same formula as $\mathbf{e} = \{\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\}\mathbf{Y}$ for ordinary least squares, but with $\hat{\mathbf{P}}^{-1}\mathbf{X}$ and $\hat{\mathbf{P}}^{-1}\mathbf{Y}$ replacing $\mathbf{X}$ and $\mathbf{Y}$ respectively.)

9.3. AN EXAMPLE OF WEIGHTED LEAST SQUARES

This is an extremely simple example but an interesting one. Suppose we wish to fit the model

$$E(Y) = \beta X.$$

Let us suppose that

$$\mathbf{V}\sigma^2 = \mathbf{V}(\mathbf{Y}) = \begin{bmatrix} 1/w_1 & & & \\ & 1/w_2 & & \mathbf{0} \\ & & \ddots & \\ \mathbf{0} & & & 1/w_n \end{bmatrix}\sigma^2,$$

where the w's are *weights* to be specified. This means that

$$\mathbf{V}^{-1} = \begin{bmatrix} w_1 & & & \\ & w_2 & & \mathbf{0} \\ & & \ddots & \\ \mathbf{0} & & & w_n \end{bmatrix}.$$

Applying the general results above we find, after reduction,

$$b = \frac{\Sigma\, w_i X_i Y_i}{\Sigma\, w_i X_i^2},$$

where all summations are from $i = 1, 2, ..., n$.

Case 1. Suppose $\sigma_i^2 = V(Y_i) = kX_i$; that is, the variance of Y_i is proportional to the size of the corresponding X_i. Then $w_i = \sigma^2/kX_i$. Hence

$$b = \frac{\Sigma\, Y_i}{\Sigma\, X_i} = \frac{\overline{Y}}{\overline{X}}.$$

Thus if the variance of Y_i is proportional to X_i, the best estimate of the regression coefficient is the mean of the Y_i divided by the mean of the X_i. In addition,

$$V(b) = \frac{\sigma^2}{\Sigma\, w_i X_i^2} = \frac{k}{\Sigma\, X_i}.$$

Case 2. Suppose $\sigma_i^2 = V(Y_i) = kX_i^2$; that is, the variance of Y_i is proportional to the square of the corresponding X_i. Then $w_i = \sigma^2/kX_i^2$. Hence

$$\begin{aligned} b &= \frac{\Sigma\, (Y_i/X_i)}{\Sigma\, 1} \\ &= \frac{\Sigma\, (Y_i/X_i)}{n}. \end{aligned}$$

Thus if the variance of the Y_i is proportional to X_i^2, the best estimate of the regression coefficient is the average of the n slopes obtained one from each pair of observations Y_i/X_i. Also,

$$V(b) = \frac{\sigma^2}{\Sigma\, w_i X_i^2} = \frac{k}{n}.$$

Note: Fitting a straight line through the origin $(X, Y) = (0, 0)$ represents a very strong assumption, which, in general, is not justified. Even when the model is "known" to pass through the origin (as would be the case, for example, if X = speed of car, Y = stopping distance) it does not mean that a straight line fit though the origin is necessarily appropriate. It may be that the available data can be fitted by a straight line not through the origin but that, if more data were available, a higher-order model that did pass through the origin would provide a proper fit. Usually it is best to put an intercept term β_0 in the model and to check on the size of the estimate b_0.

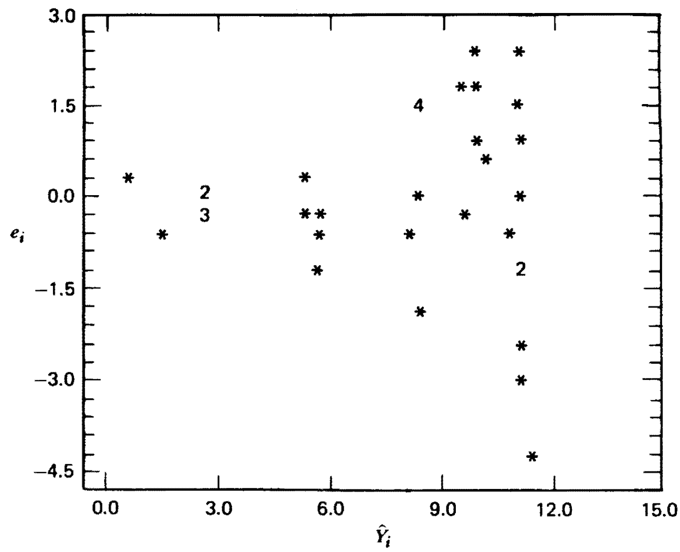
9.4 A NUMERICAL EXAMPLE OF WEIGHTED LEAST SQUARES

The data in Table 9.1, which have been rearranged in an order convenient for purposes of analysis, consist of 35 observations (X_i, Y_i) with a number of sets that are either exact repeats at the same X-value or approximate repeats. These are indicated by the groupings. A fit of the data by (ordinary) least squares produces the fitted model

TABLE 9.1. Data for Weighted Least Squares Example

X	Y	$\hat{w}_i$
1.15	0.99	1.24028
1.90	0.98	2.18224
3.00	2.60	7.84930
3.00	2.67	7.84930
3.00	2.66	7.84930
3.00	2.78	7.84930
3.00	2.80	7.84930
5.34	5.92	7.43652
5.38	5.35	6.99309
5.40	4.33	6.78574
5.40	4.89	6.78574
5.45	5.21	6.30514
7.70	7.68	0.89204
7.80	9.81	0.84420
7.81	6.52	0.83963
7.85	9.71	0.82171
7.87	9.82	0.81296
7.91	9.81	0.79588
7.94	8.50	0.78342
9.03	9.47	0.47385
9.07	11.45	0.46621
9.11	12.14	0.45878
9.14	11.50	0.45327
9.16	10.65	0.44968
9.37	10.64	0.41435
10.17	9.78	0.31182
10.18	12.39	0.31079
10.22	11.03	0.30672
10.22	8.00	0.30672
10.22	11.90	0.30672
10.18	8.68	0.31079
10.50	7.25	0.28033
10.23	13.46	0.30571
10.03	10.19	0.32680
10.23	9.93	0.30571

Source: Wanda M. Hinshaw.



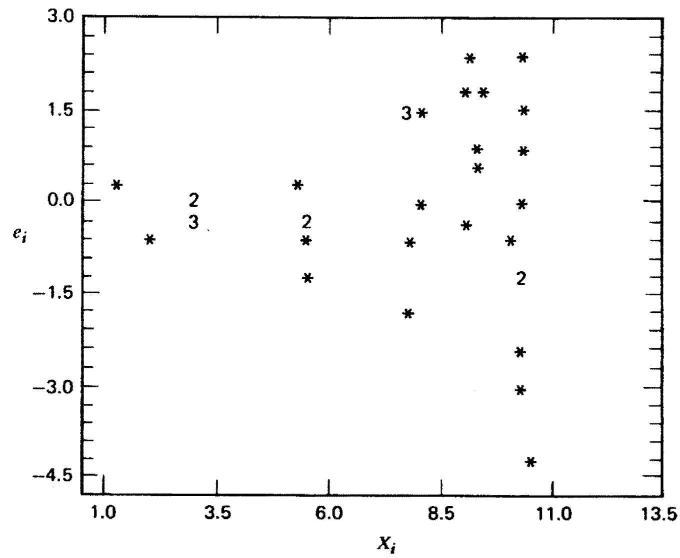


Figure 9.1. Residuals plots, unweighted least squares. (Two indistinguishable points are shown as a 2, and so on.)

$\hat{Y} = -0.5790 + 1.1354X$ and the residuals plots in Figure 9.1. A clear indication that the observations have unequal variances is seen. The overall plot of residuals (not shown) is somewhat skewed toward negative values, also. None of the usual (ordinary) least squares analyses are appropriate and it seems sensible to apply generalized least squares.

We assume (until contrary indications appear) that the Y_i are independent so that $\mathbf{V}$ has the diagonal pattern with different variances given earlier. We now need to obtain information on the variance pattern. For each of the sets of repeats or near repeats we evaluate the average X-value, X_j, say, and the pure error mean square s_{ej}^2. These are:

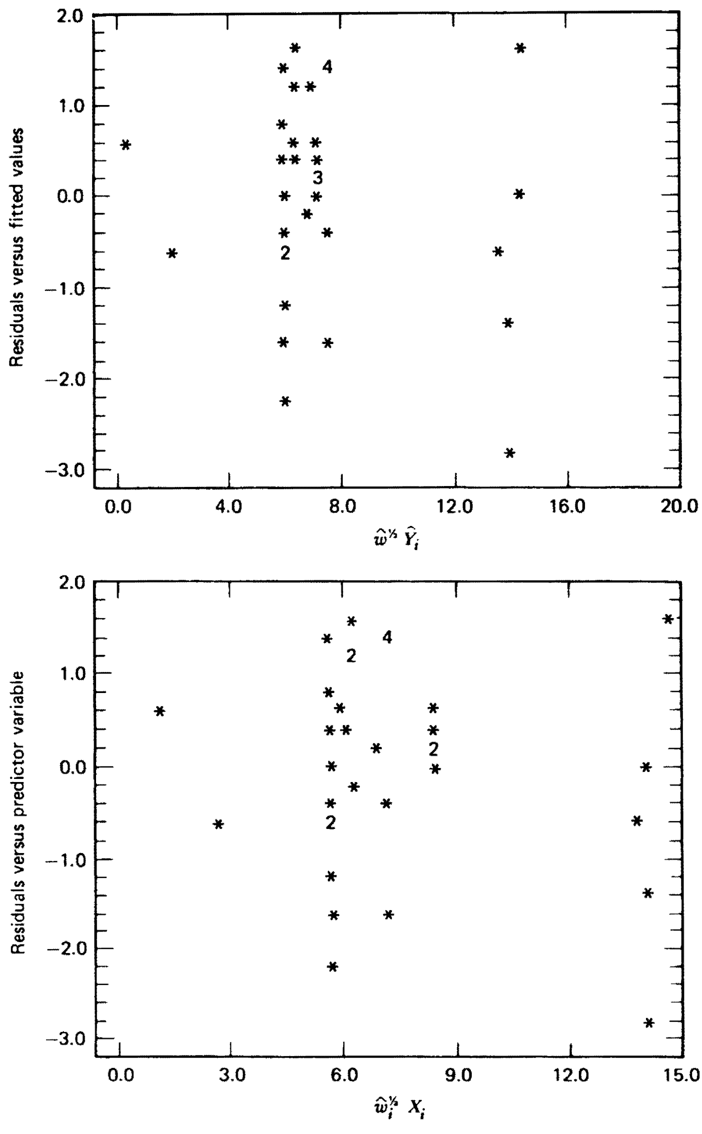


Figure 9.2. Residuals plots, weighted least squares.

$\overline{X}_j$	3.0	5.4	7.8	9.1	10.2
s_{ej}^2	0.0072	0.3440	1.7404	0.8683	3.8964

A plot of these suggests a quadratic relationship, which we estimate by least squares as

$$\hat{s}_e^2 = 1.5329 - 0.7334\overline{X} + 0.0883\overline{X}^2.$$

We can now substitute each individual X_i into this equation, estimate s_{ej}^2, $i = 1, 2, \ldots, 35$, and invert these values to give the estimated weights $\hat{w}_i$ shown in the table. The matrix **P** of our text is diagonal with entries $\hat{w}_i^{-1/2}$. Using these weights leads to the weighted least squares prediction equation $\hat{Y} = -0.8891 + 1.1648X$ and an analysis of variance table as follows:

Source	df	SS	MS
$b_1 \mid b_0$	1	496.96	496.96
Residual	33	42.66	1.29
Total, corrected	34	539.62	

The appropriate "observations" and "fitted values" are now $\hat{w}^{1/2}Y_i$ and $\hat{w}_i^{1/2}\hat{Y}_i$ and the "residuals" to be examined are $\hat{w}_i^{1/2}(Y_i - \hat{Y}_i)$, notice. An overall plot of residuals still shows some skewness but the pattern is slightly better behaved. The residuals plots in Figure 9.2 reveal that the vertical spread of residuals is now roughly the same at the two main levels of the transformed response. (At lower levels there are only two observations so that there is not much of an estimate of the spread there.) The employment of weighted least squares here appears to be justified and useful.

A weighted least squares program exists in most computing systems, but some do not have a generalized least squares program.

9.5 RESTRICTED LEAST SQUARES

For least squares involving restrictions on the parameters see, for example, Waterman (1974) and Judge and Takayama (1966). If the restrictions are of the equality form $\mathbf{C}\boldsymbol{\beta} = \mathbf{d}$, we can use the method of Lagrange's undetermined multipliers (see Appendix 9A) and minimize the Lagrangean function

$$F = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\lambda}'(\mathbf{d} - \mathbf{C}\boldsymbol{\beta}) \tag{9.5.1}$$

with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$. The solution for $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = \mathbf{b} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'[\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}']^{-1}(\mathbf{d} - \mathbf{C}\mathbf{b}), \tag{9.5.2}$$

where $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ is the usual unrestricted estimator. See also Chapters 20 and 21 for geometrical aspects.

Note that, if $\mathbf{d} = \mathbf{0}$, so that the restriction is $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$, we are back in the context of Section 9.1. Even if $\mathbf{d} \neq \mathbf{0}$, we can always substitute back into the model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ for some parameters using the restrictions $\mathbf{C}\boldsymbol{\beta} = \mathbf{d}$, to obtain a solution in terms of fewer parameters. The solution (9.5.2) is more elegant in that it retains all the original parameters and also ensures that $\mathbf{C}\hat{\boldsymbol{\beta}} = \mathbf{d}$; the latter is obvious from premultiplying Eq. (9.5.2) by $\mathbf{C}$. However, it provides the same predicted values that we would obtain by using the substitution method.

9.6. INVERSE REGRESSION (MULTIPLE PREDICTOR CASE)

Given a fitted regression equation $\hat{Y} = b_0 + b_1X_1 + b_2X_2 + \cdots + b_kX_k$ and a true mean value of Y, say, Y_0, we require a "fiducial region" for the point $(X_1, X_2, \ldots, X_k)$. Extending Eq. (3.2.8) we obtain the following equation satisfied by the boundaries of the required region:

$$\{-Y_0 + b_0 + b_1X_1 + b_2X_2 + \cdots + b_kX_k\}^2$$

$$= t^2s^2\left\{(1, X_1, X_2, \ldots, X_k)(\mathbf{X}'\mathbf{X})^{-1}\begin{bmatrix}1\\X_1\\X_2\\\vdots\\X_k\end{bmatrix}\right\}. \tag{9.6.1}$$

This is a hyperbolic surface. Figure 9.3 shows the $k = 2$ case. When Y_0 is the mean of q observations, insert "$1/q$ +" inside the curly braces on the right-hand side of Eq. (9.6.1). If $\hat{Y}$ is a polynomial, not a plane, the obvious adjustments must be made on both sides of Eq. (9.6.1).

Note: The method indicated above can be applied to other types of problems. For example, the maxima and minima of $\hat{Y} = b_0 = b_1X + b_2X^2 + b_3X^3 + b_4X^4$ are at the roots of $f \equiv b_1 + 2b_2X + 3b_3X^2 + 4b_4X^3 = 0$. Fiducial limits for the roots can be evaluated from the equation

$$f^2 = t^2s^2\{V(f)/\sigma^2\}, \tag{9.6.2}$$

where $V(f)$ denotes the variance of the function f, which has a factor g^2 in it. For a fuller account, including possible problems with imaginary roots, see Williams (1959, pp. 108–109 and 114–116). See also Box and Hunter (1954).

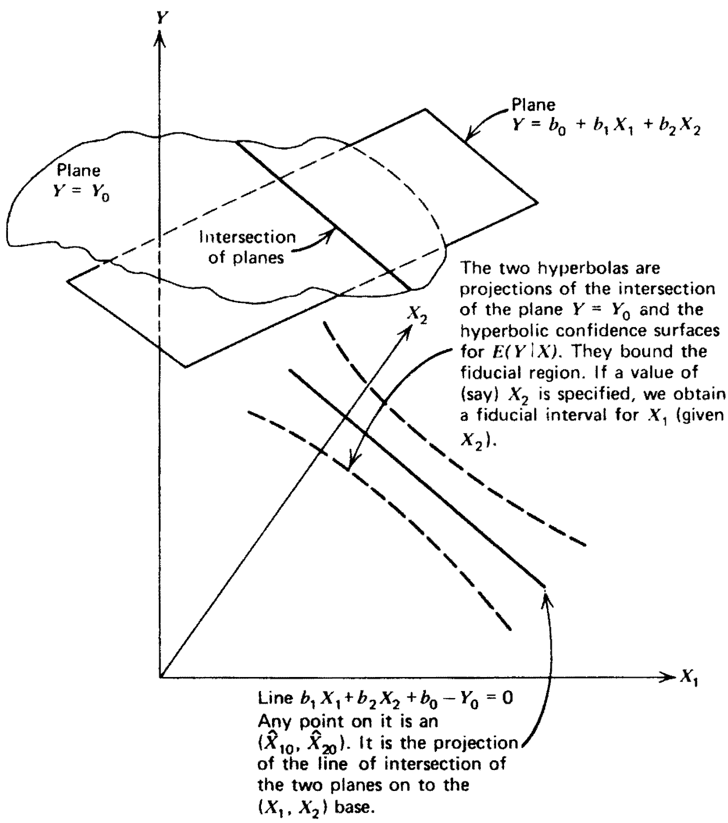


Figure 9.3. Inverse regression for two predictors.

9.7. PLANAR REGRESSION WHEN ALL THE VARIABLES ARE SUBJECT TO ERROR

We briefly describe an extension of the work of Section 3.4 for more than one X. Suppose we have observations $(Y_i, X_{1i}, X_{2i}, \ldots, X_{ki})$, $i = 1, 2, \ldots, n$, with all variables subject to random errors. Consider the quantity

$$D_{Yi}^2 = \left(\beta_0 + \sum_{j=1}^{k} \beta_j X_{ji} - Y_i\right)^2, \tag{9.7.1}$$

the squared deviation of the ith point measured in the Y direction. Similar squared deviations in other directions are defined by, for $\ell = 1, 2, \ldots, k$,

$$D_{\ell i}^2 = \left(\beta_0 + \sum_{j=1}^{k} \beta_j X_{ji} - Y_i\right)^2 / \beta_\ell^2. \tag{9.7.2}$$

The geometric mean of the D^2 values,

$$G_i = (D_{Yi}^2 D_{1i}^2 D_{2i}^2 \cdots D_{ki}^2)^{1/(k+1)}, \tag{9.7.3}$$

can also be regarded as

$$G_i = V_i^{2/(k+1)}, \tag{9.7.4}$$

where V_i is the volume created by drawing, from the ith data point $(Y_i, X_{1i}, X_{2i}, \ldots X_{ki})$, lines parallel to the $Y, X_1, X_2, \ldots, X_k$ axes to the plane $Y = \beta_0 + \sum_{j=1}^{k} \beta_j X_j$.

For $k = 1$, V_i is the area of the ith right-angled triangle in Figure 3.5, so that the general G_i is an extension of this concept. The criterion

$$L_G^k = \sum_{i=1}^{n} G_i \tag{9.7.5}$$

can now be minimized to obtain estimates for the general dimension case. For additional details, see Draper and Yang (1997).

APPENDIX 9A. LAGRANGE'S UNDETERMINED MULTIPLIERS

Notation

Because the method of Lagrange's undetermined multipliers has wide applicability, we have chosen to adopt a fairly "neutral" notation $\theta_1, \theta_2, \ldots, \theta_m$ for the variables involved in the functions f and g_j below. When the method is applied as in Section 9.5, the θ's would be *all* the parameters in the $\boldsymbol{\beta}$ vector. For the ridge regression application in Chapter 17, the θ's would be all the regression β's *except* β_0. In other applications, the θ's might be predictor variables, that is, X's.

Basic Method

Suppose we wish to obtain the stationary or turning values of a function $f(\theta_1, \theta_2, \ldots, \theta_m)$ of m variables $\theta_1, \theta_2, \ldots, \theta_m$, subject to restrictions on the θ_i such as

$$g_j(\theta_1, \theta_2, \ldots, \theta_m) = 0 \qquad (j = 1, 2, \ldots, q).$$

Form the function

$$F = f - \sum_{j=1}^{q} \lambda_j g_j, \tag{9A.1}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_q$ are unknowns. Differentiate Eq. (9A.1) partially with respect to each θ_i and set the results equal to zero. This will provide the m equations

$$\frac{\partial F}{\partial \theta_i} \equiv \frac{\partial f}{\partial \theta_i} - \sum_{j=1}^{q} \lambda_j \frac{\partial g_j}{\partial \theta_i} = 0 \qquad (i = 1, 2, \ldots, m). \tag{9A.2}$$

These m equations, with the additional q equations

$$g_j = 0 \qquad (j = 1, 2, \ldots, q), \tag{9A.3}$$

provide $(q + m)$ equations that can be solved for the $(q + m)$ unknowns $\theta_1, \theta_2, \ldots, \theta_m, \lambda_1, \lambda_2, \ldots, \lambda_q$. Often the quantities λ_j are eliminated and not actually found; for this reason the words "undetermined multipliers" are used to describe them. In some cases, however, the solutions for $\theta_1, \theta_2, \ldots, \theta_m$ are easier to obtain if the λ_j are evaluated first; in other cases, it may be easier to specify values of λ_j in Eqs. (9A.2) and regard other quantities in Eqs. (9A.3) as unknowns, in their place.

Is the Solution a Maximum or Minimum?

Suppose now that $(\theta_1, \theta_2, \ldots, \theta_m) = (a_1, a_2, \ldots, a_m)$ is a solution of Eqs. (9A.2) and (9A.3) after elimination of λ_j. Let

$$\mathbf{M}(\theta) = \mathbf{M}(\theta_1, \theta_2, \ldots, \theta_m) = \begin{bmatrix} \dfrac{\partial^2 F}{\partial \theta_1^2} & \dfrac{\partial^2 F}{\partial \theta_1 \partial \theta_2} & \cdots & \dfrac{\partial^2 F}{\partial \theta_1 \partial \theta_m} \\ \dfrac{\partial^2 F}{\partial \theta_2 \partial \theta_1} & \dfrac{\partial^2 F}{\partial \theta_2^2} & \cdots & \dfrac{\partial^2 F}{\partial \theta_2 \partial \theta_m} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial^2 F}{\partial \theta_m \partial \theta_1} & \dfrac{\partial^2 F}{\partial \theta_m \partial \theta_2} & \cdots & \dfrac{\partial^2 F}{\partial \theta_m^2} \end{bmatrix} \tag{9A.4}$$

be the matrix of second-order partial derivatives. Let $\mathbf{M}(a_1, a_2, \ldots, a_m) = \mathbf{M}(\mathbf{a})$ be the resulting matrix after the solution $\mathbf{a}' = (a_1, a_2, \ldots, a_m)$ has been substituted into Eq. (9A.4). Then if $\mathbf{M}(\mathbf{a})$ is

1. positive definite, that is, $\mathbf{u}'\mathbf{M}\mathbf{u} > 0$, for all $\mathbf{u}$,
2. negative definite, that is, $\mathbf{u}'\mathbf{M}\mathbf{u} < 0$, for all $\mathbf{u}$,

where $\mathbf{u}' = (u_1, u_2, \ldots, u_m)$ is any $1 \times m$ real vector, the function $f(\theta_1, \theta_2, \ldots, \theta_m)$ achieves

1. a local minimum at $\boldsymbol{\theta} = \mathbf{a}$,
2. a local maximum at $\boldsymbol{\theta} = \mathbf{a}$,

respectively. For, if we expand F about $\mathbf{a}$ as a Taylor series of partial derivatives, remembering that all first partial derivatives of F are zero at $\boldsymbol{\theta} = \mathbf{a}$, we see that

$$F(\mathbf{a} + \mathbf{h}) - F(\mathbf{a}) = \tfrac{1}{2}\mathbf{h}'\mathbf{M}(\mathbf{a})\mathbf{h} + O(h^3),$$

where $\mathbf{h}$ represents a vector of small increments h_i all of the same order and $O(h^3)$

represents a remainder of third order in such increments. Thus, to order h^2, if $\mathbf{M}(\mathbf{a})$ is positive definite, for example, then

$$F(\mathbf{a}+\mathbf{h}) > F(\mathbf{a}), \qquad \text{for } all \text{ small } \mathbf{h}.$$

If $\mathbf{h}$ varies only in such a way that the restrictions are still satisfied, this implies that

$$f(\mathbf{a}+\mathbf{h}) > f(\mathbf{a}),$$

that is, $f(\mathbf{a})$ is, locally, a minimum, subject to the restrictions holding. As we can see from this discussion, it might happen that

$$F(\mathbf{a}+\mathbf{h}) \not> F(\mathbf{a}), \qquad \text{for all small } \mathbf{h}$$

but

$$f(\mathbf{a}+\mathbf{h}) > f(\mathbf{a}),$$

for all $\mathbf{h}$ that satisfy the restrictions. Thus "$\mathbf{M}(\mathbf{a})$ is positive definite" is sufficient, but not necessary, for a local restricted minimum of f at $\theta = \mathbf{a}$. Similar remarks apply to the negative definite case. If $\mathbf{M}(\mathbf{a})$ is indefinite, further investigation of the function near the point $\mathbf{a}$ is required to determine what sort of stationary point has been obtained.

EXERCISES FOR CHAPTER 9

A. Consider the model $Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \epsilon$. If it is suggested to you that the two variables $Z_1 = X_1 + X_3$ and $Z_2 = X_2 + X_4$ might be adequate to represent the data, what hypothesis, in the form $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$, would you need to test? (Give the form of $\mathbf{C}$.)

B. For the data $(X_1, X_2, Y) = (-1, -1, 7.2), (-1, 0, 8.1), (0, 0, 9.8), (1, 0, 12.3), (1, 1, 12.9)$, the least squares fit is $\hat{Y} = 10.6 + 2.10X_1 + 0.75X_2$, and the residual sum of squares is 0.107 (2 df). Test the null hypothesis $H_0: \beta_1 = 2\beta_2$ versus H_1: not so.

C. Look at the Hald data in Appendix 15A. Fit $Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \epsilon$ (which is done in Appendix 15A) and test $H_0: \beta_1 = \beta_3, \beta_2 = \beta_4$, versus the alternative not so. Is this a reasonable hypothesis to test?

D. Consider the data of Table 2.1. Suppose you are told that the 23rd observation has variance $4\sigma^2$ rather than σ^2. Refit the equation using weighted least squares with $\mathbf{V}^{-1} = (1, 1, \ldots, 1, 0.25)$.

E. Repeat Exercise D but with $16\sigma^2$ for the variance of the last observation. What changes do you observe?

F. (*Source:* J. A. John.) An experimenter tells you he wishes to fit the model $Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \epsilon$ by least squares, subject to the restriction that $\beta_1 = 1$. He asks specifically if he can just fit $Y - X_1 = \beta_0 + \beta_2X_2 + \epsilon$ by least squares to get what he wants. Can he? (Yes.)

G. (*Source:* T. J. Mitchell.) An experimenter wishes to fit the quadratic model $Y = \beta_0 + \beta_1X + \beta_{11}X^2 + \epsilon$. He "knows" (he says) that the response at $X_1 = 1$ is 10 so that, ignoring the error term in the model, $10 = \beta_0 + \beta_1 + \beta_{11}$. He then substitutes for $\beta_0 = 10 - \beta_1 - \beta_{11}$ in the first model to give $Y - 10 = \beta_1Z_1 + \beta_{11}Z_2 + \epsilon$, where $Z_1 = X - 1$ and $Z_2 = X^2 - 1$. He next fits this second model by least squares to provide b_1 and b_{11}, determines $b_0 = 10 - b_1 - b_{11}$, and announces he has obtained the least squares solution for the first model, subject to the restriction that the response at $X = 1$ is 10. Is he correct? (Yes.)

H. (*Source:* S. C. Piper.) Suppose we wish to fit the model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \epsilon$$

by least squares, but it is true that

$$\beta_1 + \beta_2 + \beta_3 + \beta_4 = C_1 \text{ (known constant)}$$

and

$$\beta_5 + \beta_6 + \beta_7 = C_2 \text{ (known constant)}.$$

Suppose we substitute for (say) β_4 and β_7 in the original model using the restrictions and then fit the resulting model

$$Y - C_1 X_4 - C_2 X_7 = \beta_0 + \beta_1(X_1 - X_4) + \beta_2(X_2 - X_4) + \beta_3(X_3 - X_4) + \beta_5(X_5 - X_7) + \beta_6(X_6 - X_7) + \epsilon$$

by least squares. Will this solution be correct? (Yes)

I. (Generalized restricted least squares.) Use the method of Lagrange's undetermined multipliers to show that, for a generalized least squares problem in which Eq. (9.5.1) is replaced by

$$F = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\lambda}'(\mathbf{d} - \mathbf{C}\boldsymbol{\beta}),$$

the solution for $\boldsymbol{\beta}$ replacing (9.5.2) is

$$\hat{\boldsymbol{\beta}} = \mathbf{b}_G + (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{C}'[\mathbf{C}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{C}']^{-1}(\mathbf{d} - \mathbf{C}\mathbf{b}_G)$$

where $\mathbf{b}_G = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}$ is the unrestricted generalized least squares estimator.