

## Chapter 6

### Regression Analysis Under Linear Restrictions and Preliminary Test Estimation

One of the basic objectives in any statistical modeling is to find good estimators of the parameters. In the context of the multiple linear regression model  $y = X\beta + \varepsilon$ , the ordinary least squares estimator  $b = (X'X)^{-1}X'y$  is the best linear unbiased estimator of  $\beta$ . Several approaches have been attempted in the literature to improve further the OLSE. One approach to improve the estimators is the use of extraneous information or prior information. In applied work, such prior information may be available about the regression coefficients. For example, in economics, the constant returns to scale imply that the exponents in a Cobb-Douglas production function should sum to unity. In another example, the absence of money illusion on the part of consumers implies that the sum of money income and price elasticities in a demand function should be zero. These types of constraints or the prior information may be available from

- (i) some theoretical considerations.
- (ii) past experience of the experimenter.
- (iii) empirical investigations.
- (iv) some extraneous sources etc.

To utilize such information in improving the estimation of regression coefficients, it can be expressed in the form of

- (i) exact linear restrictions
- (ii) stochastic linear restrictions
- (iii) inequality restrictions.

We consider the use of prior information in the form of exact and stochastic linear restrictions in the model  $y = X\beta + \varepsilon$  where  $y$  is a  $(n \times 1)$  vector of observations on study variable,  $X$  is a  $(n \times k)$  matrix of observations on explanatory variables  $X_1, X_2, \dots, X_k$ ,  $\beta$  is a  $(k \times 1)$  vector of regression coefficients and  $\varepsilon$  is a  $(n \times 1)$  vector of disturbance terms.

## Exact linear restrictions:

Suppose the prior information binding the regression coefficients is available from some extraneous sources which can be expressed in the form of exact linear restrictions as

$$r = R\beta$$

where  $r$  is a  $(q \times 1)$  vector and  $R$  is a  $(q \times k)$  matrix with  $\text{rank}(R) = q$  ( $q < k$ ). The elements in  $r$  and  $R$  are known.

Some examples of exact linear restriction  $r = R\beta$  are as follows:

- (i) If there are two restrictions with  $k = 6$  like

$$\begin{aligned}\beta_2 &= \beta_4 \\ \beta_3 + 2\beta_4 + \beta_5 &= 1\end{aligned}$$

then

$$r = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, R = \begin{bmatrix} 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 2 & 1 & 0 \end{bmatrix}.$$

- (ii) If  $k = 3$  and suppose  $\beta_2 = 3$ , then

$$r = [3], R = [0 \ 1 \ 0]$$

- (iii) If  $k = 3$  and suppose  $\beta_1 : \beta_2 : \beta_3 :: ab : b : 1$

$$\text{then } r = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, R = \begin{bmatrix} 1 & -a & 0 \\ 0 & 1 & -b \\ 1 & 0 & -ab \end{bmatrix}.$$

The ordinary least squares estimator  $b = (X'X)^{-1}X'y$  does not use the prior information. It does not obey the restrictions in the sense that  $r \neq Rb$ . So the issue is how to use the sample information and prior information together in finding an improved estimator of  $\beta$ .

## Restricted least squares estimation

The restricted least squares estimation method enables the use of sample information and prior information simultaneously. In this method, choose  $\beta$  such that the error sum of squares is minimized subject to linear restrictions  $r = R\beta$ . This can be achieved using the Lagrangian multiplier technique. Define the Lagrangian function

$$S(\beta, \lambda) = (y - X\beta)'(y - X\beta) - 2\lambda'(R\beta - r)$$

where  $\lambda$  is a  $(k \times 1)$  vector of the Lagrangian multiplier.

Using the result that if  $a$  and  $b$  are vectors and  $A$  is a suitably defined matrix, then

$$\frac{\partial}{\partial a} a' A a = (A + A')a$$

$$\frac{\partial}{\partial a} a' b = b,$$

we have

$$\frac{\partial S(\beta, \lambda)}{\partial \beta} = 2X'X\beta - 2X'y - 2R'\lambda = 0 \quad (*)$$

$$\frac{\partial S(\beta, \lambda)}{\partial \lambda} = R\beta - r = 0.$$

Pre-multiplying equation (\*) by  $R(X'X)^{-1}$ , we have

$$2R\beta - 2R(X'X)^{-1}X'y - 2R(X'X)^{-1}R'\lambda' = 0$$

$$\text{or } R\beta - Rb - R(X'X)^{-1}R'\lambda' = 0$$

$$\Rightarrow \lambda' = -[R(X'X)^{-1}R']^{-1}(Rb - r)$$

using  $R(X'X)^{-1}R' > 0$ .

Substituting  $\lambda$  in equation (\*), we get

$$2X'X\beta - 2X'y + 2R'[R(X'X)^{-1}R']^{-1}(Rb - r) = 0$$

$$\text{or } X'X\beta = X'y - R'[R(X'X)^{-1}R']^{-1}(Rb - r).$$

Pre-multiplying by  $(X'X)^{-1}$  yields

$$\begin{aligned} \hat{\beta}_R &= (X'X)^{-1}X'y + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(r - Rb) \\ &= b - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(Rb - r). \end{aligned}$$

This estimation is termed as **restricted regression estimator** of  $\beta$ .

## Properties of restricted regression estimator

1. The restricted regression estimator  $\hat{\beta}_R$  obeys the exact restrictions, i.e.,  $r = R\hat{\beta}_R$ . To verify this, consider

$$\begin{aligned} R\hat{\beta}_R &= R \left[ b + (X'X)^{-1} R' \{ R(X'X)^{-1} R' \}^{-1} (r - Rb) \right] \\ &= Rb + r - Rb \\ &= r. \end{aligned}$$

## 2. Unbiasedness

The estimation error of  $\hat{\beta}_R$  is

$$\begin{aligned} \hat{\beta}_R - \beta &= (b - \beta) + (X'X)^{-1} R' \left[ R(X'X)^{-1} R' \right]^{-1} (R\beta - Rb) \\ &= \left[ I - (X'X)^{-1} R' \{ R(X'X)^{-1} R' \}^{-1} R \right] (b - \beta) \\ &= D(b - \beta) \end{aligned}$$

where

$$D = I - (X'X)^{-1} R' \left[ R(X'X)^{-1} R' \right]^{-1} R.$$

Thus

$$\begin{aligned} E(\hat{\beta}_R - \beta) &= DE(b - \beta) \\ &= 0 \end{aligned}$$

implying that  $\hat{\beta}_R$  is an unbiased estimator of  $\beta$ .

## 3. Covariance matrix

The covariance matrix of  $\hat{\beta}_R$  is

$$\begin{aligned} V(\hat{\beta}_R) &= E(\hat{\beta}_R - \beta)(\hat{\beta}_R - \beta)' \\ &= DE(b - \beta)(b - \beta)'D' \\ &= DV(b)D' \\ &= \sigma^2 D(X'X)^{-1} D' \\ &= \sigma^2 (X'X)^{-1} - \sigma^2 (X'X)^{-1} R' \left[ R(X'X)^{-1} R' \right]^{-1} R(X'X)^{-1} \end{aligned}$$

which can be obtained as follows:

Consider

$$\begin{aligned}
D(X'X)^{-1} &= (X'X)^{-1} - (X'X)^{-1} R' [R(X'X)^{-1} R']^{-1} R (X'X)^{-1} \\
D(X'X)^{-1} D' &= \left[ (X'X)^{-1} - (X'X)^{-1} R' \{R(X'X)^{-1} R'\}^{-1} R (X'X)^{-1} \right] \left[ I - (X'X)^{-1} R' \{R(X'X)^{-1} R'\}^{-1} R \right]' \\
&= (X'X)^{-1} - (X'X)^{-1} R' [R(X'X)^{-1} R']^{-1} R (X'X)^{-1} - (X'X)^{-1} R' [R(X'X)^{-1} R']^{-1} R (X'X)^{-1} \\
&\quad + (X'X)^{-1} R' [R(X'X)^{-1} R']^{-1} R (X'X)^{-1} R' [R(X'X)^{-1} R']^{-1} R (X'X)^{-1} \\
&= (X'X)^{-1} - (X'X)^{-1} R' [R(X'X)^{-1} R']^{-1} R (X'X)^{-1}.
\end{aligned}$$

### Maximum likelihood estimation under exact restrictions:

Assuming  $\varepsilon \sim N(0, \sigma^2 I)$ , the maximum likelihood estimator of  $\beta$  and  $\sigma^2$  can also be derived such that it follows  $r = R\beta$ . The Lagrangian function as per the maximum likelihood procedure can be written as

$$L(\beta, \sigma^2, \lambda) = \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp \left[ -\frac{1}{2} \left\{ \frac{(y - X\beta)'(y - X\beta)}{\sigma^2} - \lambda'(R\beta - r) \right\} \right]$$

where  $\lambda$  is a  $(q \times 1)$  vector of Lagrangian multipliers. The normal equations are obtained by partially differentiating the log-likelihood function with respect to  $\beta, \sigma^2$  and  $\lambda$  and equated to zero as

$$\frac{\partial \ln L(\beta, \sigma^2, \lambda)}{\partial \beta} = -\frac{1}{\sigma^2} (X'X\beta - X'y) + 2R'\lambda = 0 \quad (1)$$

$$\frac{\partial \ln L(\beta, \sigma^2, \lambda)}{\partial \lambda} = 2(R\beta - r) = 0 \quad (2)$$

$$\frac{\partial \ln L(\beta, \sigma^2, \lambda)}{\partial \sigma^2} = -\frac{2n}{\sigma^2} + \frac{2(y - X\beta)'(y - X\beta)}{\sigma^4} = 0. \quad (3)$$

Let  $\tilde{\beta}_R, \tilde{\sigma}_R^2$  and  $\tilde{\lambda}$  denote the maximum likelihood estimators of  $\beta, \sigma^2$  and  $\lambda$  respectively which are obtained by solving equations (1), (2) and (3) as follows:

From equation (1), we get optimal  $\lambda$  as

$$\tilde{\lambda} = \frac{[R(X'X)^{-1} R']^{-1} (r - R\tilde{\beta})}{\tilde{\sigma}^2}.$$

Substituting  $\tilde{\lambda}$  in equation (1) gives

$$\tilde{\beta}_R = \tilde{\beta} + (X'X)^{-1} R' [R(X'X)^{-1} R']^{-1} (r - R\tilde{\beta})$$

where  $\tilde{\beta} = (X'X)^{-1} X'y$  is the maximum likelihood estimator of  $\beta$  without restrictions. From equation (3), we get

$$\tilde{\sigma}_R^2 = \frac{(y - X\tilde{\beta})'(y - X\tilde{\beta})}{n}.$$

The Hessian matrix of second-order partial derivatives of  $\beta$  and  $\sigma^2$  is positive definite at  $\beta = \tilde{\beta}_R$  and  $\sigma^2 = \tilde{\sigma}_R^2$ .

The restricted least squares and restricted maximum likelihood estimators of  $\beta$  are the same whereas they are different for  $\sigma^2$ .

## Test of hypothesis

It is important to test the hypothesis

$$H_0 : r = R\beta$$

$$H_1 : r \neq R\beta$$

before using it in the estimation procedure.

The construction of the test statistic for this hypothesis is detailed in the module on a multiple linear regression model. The resulting test statistic is

$$F = \frac{\left( \frac{(r - Rb)' [R(X'X)^{-1} R']^{-1} (r - Rb)}{q} \right)}{\left( \frac{(y - Xb)'(y - Xb)}{n - k} \right)}$$

which follows a  $F$ -distribution with  $q$  and  $(n - k)$  degrees of freedom under  $H_0$ . The decision rule is to reject  $H_0$  at  $\alpha$  level of significance whenever

$$F \geq F_{1-\alpha}(q, n - k).$$

## Stochastic linear restrictions:

The exact linear restrictions assume that there is no randomness involved in the auxiliary or prior information. This assumption may not hold true in many practical situations, and some randomness may be present. The prior information in such cases can be formulated as

$$r = R\beta + V$$

where  $r$  is a  $(q \times 1)$  vector,  $R$  is a  $(q \times k)$  matrix and  $V$  is a  $(q \times 1)$  vector of random errors. The elements in  $r$  and  $R$  are known. The term  $V$  reflects the randomness involved in the prior information  $r = R\beta$ .

Assume

$$E(V) = 0$$

$$E(VV') = \psi$$

$$E(\varepsilon V') = 0.$$

where  $\psi$  is a known  $(q \times q)$  positive definite matrix and  $\varepsilon$  is the disturbance term in a multiple regression model  $y = X\beta + \varepsilon$ .

Note that  $E(r) = R\beta$ .

The possible reasons for such stochastic linear restriction are as follows:

- (i) Stochastic linear restrictions exhibit the instability of estimates. An unbiased estimate with the standard error may exhibit stability. For example, in repetitive studies, the surveys are conducted every year. Suppose the regression coefficient  $\beta_1$  remains stable for several years. Suppose its estimate is provided along with its standard error. Suppose its value remains stable around the value 0.5 with standard error 2. This information can be expressed as

$$r = \beta_1 + V_1,$$

where  $r = 0.5$ ,  $E(V_1) = 0$ ,  $E(V_1^2) = 2^2$ .

Now  $\psi$  can be formulated with this data. It is not necessary that we should have information for all regression coefficients, but we can have information on some of the regression coefficients only.

- (ii) Sometimes the restrictions are in the form of inequality. Such restrictions may arise from theoretical considerations. For example, the value of a regression coefficient may lie between 3 and 5, i.e.,  $3 \leq \beta_1 \leq 5$ , say. In another example, consider a simple linear regression model

$$y = \beta_0 + \beta_1 x + \varepsilon$$

where  $y$  denotes the consumption expenditure on food and  $x$  denotes the income. Then the marginal propensity (tendency) to consume is

$$\frac{dy}{dx} = \beta_1,$$

i.e., if salary increase by rupee one, then one is expected to spend  $\beta_1$ , amount of rupee one on food or save  $(1 - \beta_1)$  amount. We may put a bound on  $\beta$  that either one can not spend all of rupee one or nothing out of rupee one. So  $0 < \beta_1 < 1$ . This is a natural restriction arising from theoretical considerations.

These bounds can be treated as  $p$ -sigma limits, say 2-sigma limits or confidence limits. Thus

$$\begin{aligned}\mu - 2\sigma &= 0 \\ \mu + 2\sigma &= 1 \\ \Rightarrow \mu &= \frac{1}{2}, \sigma = \frac{1}{4}.\end{aligned}$$

These values can be interpreted as

$$\begin{aligned}\beta_1 + V_1 &= \frac{1}{2} \\ E(V_1^2) &= \frac{1}{16}.\end{aligned}$$

- (iii) Sometimes the truthfulness of exact linear restriction  $r = R\beta$  can be suspected and accordingly, an element of uncertainty can be introduced. For example, one may say that 95% of the restrictions hold. So some element of uncertainty prevails.

## Pure and mixed regression estimation:

Consider the multiple regression model

$$y = X\beta + \varepsilon$$

with  $n$  observations and  $k$  explanatory variables  $X_1, X_2, \dots, X_k$ . The ordinary least squares estimator of  $\beta$  is

$$b = (X'X)^{-1} X'y$$

which is termed as **pure estimator**. The pure estimator  $b$  does not satisfy the restriction  $r = R\beta + V$ . So the objective is to obtain an estimate of  $\beta$  by utilizing the stochastic restrictions such that the resulting estimator satisfies the stochastic restrictions also. In order to avoid the conflict between prior information and sample information, we can combine them as follows:



Write

$$\begin{aligned} y &= X\beta + \varepsilon & E(\varepsilon) &= 0, E(\varepsilon\varepsilon') = \sigma^2 I_n \\ r &= R\beta + V & E(V) &= 0, E(VV') = \psi, E(\varepsilon V') = 0 \end{aligned}$$

jointly as

$$\begin{bmatrix} y \\ r \end{bmatrix} = \begin{bmatrix} X \\ R \end{bmatrix} \beta + \begin{bmatrix} \varepsilon \\ V \end{bmatrix}$$

or  $a = A\beta + w$

where  $a = (y \quad r)'$ ,  $A = (X \quad R)'$ ,  $w = (\varepsilon \quad V)'$ .

Note that

$$\begin{aligned} E(w) &= \begin{bmatrix} E(\varepsilon) \\ E(V) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \Omega &= E(w w') \\ &= E \begin{bmatrix} \varepsilon\varepsilon' & \varepsilon V' \\ V\varepsilon' & VV' \end{bmatrix} \\ &= \begin{bmatrix} \sigma^2 I_n & 0 \\ 0 & \psi \end{bmatrix}. \end{aligned}$$

This shows that the disturbances  $w$  are non-spherical or heteroskedastic. So the application of generalized least squares estimation will yield more efficient estimator than ordinary least squares estimation. So applying generalized least squares to the model

$$a = AB + w \quad E(w) = 0, V(w) = \Omega,$$

the generalized least square estimator of  $\beta$  is given by

$$\hat{\beta}_M = (A' \Omega^{-1} A)^{-1} A' \Omega^{-1} a.$$

The explicit form of this estimator is obtained as follows:

$$\begin{aligned} A' \Omega^{-1} a &= \begin{bmatrix} X' & R' \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma^2} I_n & 0 \\ 0 & \Psi^{-1} \end{bmatrix} \begin{bmatrix} y \\ r \end{bmatrix} \\ &= \frac{1}{\sigma^2} X' y + k' \Psi^{-1} r \end{aligned}$$

$$A' \Omega^{-1} A = \begin{bmatrix} X' & R' \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma^2} I_n & 0 \\ 0 & \Psi^{-1} \end{bmatrix} \begin{bmatrix} X \\ R \end{bmatrix}$$

$$= \frac{1}{\sigma^2} X' X + R' \Psi^{-1} R.$$

Thus

$$\hat{\beta}_M = \left( \frac{1}{\sigma^2} X' X + R' \Psi^{-1} R \right)^{-1} \left( \frac{1}{\sigma^2} X' y + R' \Psi^{-1} r \right)$$

assuming  $\sigma^2$  to be unknown. This is termed as **mixed regression estimator**.

If  $\sigma^2$  is unknown, then  $\sigma^2$  can be replaced by its estimator  $\hat{\sigma}^2 = s^2 = \frac{1}{n-k} (y - Xb)'(y - Xb)$  and **feasible mixed regression estimator** of  $\beta$  is obtained as

$$\hat{\beta}_f = \left( \frac{1}{s^2} X' X + R' \Psi^{-1} R \right)^{-1} \left( \frac{1}{s^2} X' y + R' \Psi^{-1} r \right).$$

This is also termed as **estimated** or **operationalized generalized least squares estimator**.

## Properties of mixed regression estimator:

### (i) Unbiasedness:

The estimation error of  $\hat{\beta}_m$  is

$$\begin{aligned} \hat{\beta}_M - \beta &= (A' \Omega A)^{-1} A' \Omega^{-1} a - \beta \\ &= (A' \Omega^{-1} A)^{-1} A' \Omega^{-1} (AB + w) - \beta \\ &= (A' \Omega^{-1} A)^{-1} A' \Omega^{-1} w. \\ E(\hat{\beta}_M - \beta) &= (A' \Omega^{-1} A)^{-1} A' \Omega^{-1} E(w) \\ &= 0. \end{aligned}$$

So mixed regression estimator provides an unbiased estimator of  $\beta$ . Note that the pure regression

$b = (X' X)^{-1} X' y$  estimator is also an unbiased estimator of  $\beta$ .

## (ii) Covariance matrix

The covariance matrix of  $\hat{\beta}_M$  is

$$\begin{aligned} V(\hat{\beta}_M) &= E(\hat{\beta}_M - \beta)(\hat{\beta}_M - \beta)' \\ &= (A' \Omega^{-1} A)^{-1} A' \Omega^{-1} E(VV') \Omega^{-1} A (A' \Omega^{-1} A)^{-1} \\ &= (A' \Omega^{-1} A)^{-1} \\ &= \left( \frac{1}{\sigma^2} X' X + R' \Omega^{-1} R \right)^{-1}. \end{aligned}$$

(iii) The estimator  $\hat{\beta}_M$  satisfies the stochastic linear restrictions in the sense that

$$\begin{aligned} r &= R\hat{\beta}_M + V \\ E(r) &= RE(\hat{\beta}_M) + E(V) \\ &= R\beta + 0 \\ &= R\beta. \end{aligned}$$

## (iv) Comparison with OLSE

We first state a result that is used further to establish the dominance of  $\hat{\beta}_M$  over  $b$ .

**Result:** The difference of matrices  $(A_1^{-1} - A_2^{-1})$  is positive definite if  $(A_2 - A_1)$  is positive definite.

Let

$$\begin{aligned} A_1 &\equiv V(b) = \sigma^2 (X' X)^{-1} \\ A_2 &\equiv V(\hat{\beta}_M) = \left( \frac{1}{\sigma^2} X' X + R' \Psi^{-1} R \right)^{-1} \end{aligned}$$

$$\begin{aligned} \text{then } A_1^{-1} - A_2^{-1} &= \frac{1}{\sigma^2} X' X + R' \Psi^{-1} R - \frac{1}{\sigma^2} X' X \\ &= R' \Psi^{-1} R \end{aligned}$$

which is a positive definite matrix. This implies that

$$A_1 - A_2 = V(b) - V(\hat{\beta}_M)$$

is a positive definite matrix. Thus  $\hat{\beta}_M$  is more efficient than  $b$  under the criterion of covariance matrices or Loewner ordering provided  $\sigma^2$  is known.

## Testing of hypothesis:

In the prior information specified by stochastic restriction  $r = R\beta + V$ , we want to test whether there is a close relation between the sample information and the prior information. The test for the compatibility of sample and prior information is tested by  $\chi^2$  – test statistic given by

$$\chi^2 = \frac{1}{\sigma^2} (r - Rb)' \left[ R(X'X)^{-1} R' + \Psi \right]^{-1} (r - Rb)$$

assuming  $\sigma^2$  is known and  $b = (X'X)^{-1} X'y$ . This follows a  $\chi^2$ -distribution with  $q$  degrees of freedom.

If  $\Psi = 0$ , then the distribution is degenerated and hence  $r$  becomes a fixed quantity. For the feasible version of mixed regression estimator

$$\hat{\beta}_f = \left( \frac{1}{s^2} X'X + R'\Psi^{-1}R \right)^{-1} \left( \frac{1}{s^2} X'y + R'\Psi^{-1}r \right),$$

the optimal properties of mixed regression estimator like linearity unbiasedness and/or minimum variance do not remain valid. So there can be situations when the incorporation of prior information may lead to a loss in efficiency. This is not a favourable situation. Under such situations, pure regression estimator is better to use. In order to know whether the use of prior information will lead to better estimator or not, the null hypothesis  $H_0 : E(r) = R\beta$  can be tested.

For testing the null hypothesis

$$H_0 : E(r) = R\beta$$

when  $\sigma^2$  is unknown, we use the  $F$  – statistic given by

$$F = \frac{\left[ (r - Rb)' \left\{ R(X'X)^{-1} R' + \Psi \right\}^{-1} (r - Rb) \right] / q}{s^2}$$

where  $s^2 = \frac{1}{n-k} (y - Xb)'(y - Xb)$  and  $F$  follows a  $F$  – distribution with  $q$  and  $(n-k)$  degrees of freedom under  $H_0$ .

## Inequality Restrictions

Sometimes the restriction on the regression parameters or equivalently the prior information about the regression parameters is available in the form of inequalities. For example,  $1 < \beta_1 < 2, 5 < \beta_3 < 6, 2 < \beta_1 + 2\beta_2 < 5$ , etc. Suppose such information is expressible in the form of  $R\beta \geq r$ . We want to estimate the regression coefficient  $\beta$  in the model  $y = X\beta + \varepsilon$  subject to constraints  $R\beta \geq r$ .

One can minimize  $(y - X\beta)'(y - X\beta)$  subject to  $R\beta \geq r$  to obtain an estimator of  $\beta$ . This can be formulated as a quadratic programming problem and can be solved using an appropriate algorithm, e.g. Simplex algorithm and a numerical solution is obtained. The advantage of this procedure is that a solution  $\hat{\beta}$  is found that fulfils the condition. The disadvantage is that the statistical properties of the estimates are not easily determined, and no general conclusions about superiority can be made.

Another option to obtain an estimator of  $\beta$  is subject to inequality constraints is to convert the inequality constraints in the form of stochastic linear restrictions e.g.,  $p$ -sigma limits, and use the framework of mixed regression estimation.

The minimax estimation can also be used to obtain the estimator of  $\beta$  under inequality constraints. The minimax estimation is based on the idea that the quadratic risk function for the estimate  $\hat{\beta}$  is not minimized over the entire parameter space but only over an area that is restricted by the prior knowledge or restrictions in relation to the estimate.

If all the restriction define a convex area, this area can be enclosed in an ellipsoid of the following form

$$B(\beta) = \{\beta : \beta' T \beta \leq k\}$$

with the origin as center point or in

$$B(\beta, \beta_0) = \{\beta : (\beta - \beta_0)' T (\beta - \beta_0) \leq k\}$$

with the center point vector  $\beta_0$  where  $k > 0$  is a given constant and  $T$  is a known  $p \times p$  matrix which is assumed to be positive definite. Here  $B$  defines a concentration ellipsoid.

First, we consider an example to understand how the inequality constraints are framed. Suppose it is known a priori that

$$a_i \leq \beta_i \leq b_i (i = 1, 2, \dots, n)$$

when  $a_i$  and  $b_i$  ( $i = 1, 2, \dots, n$ ) are known and may include  $a_i = -\infty$  and  $b_i = \infty$ . These restrictions can be written as

$$\frac{\left| \beta_i - \frac{a_i + b_i}{2} \right|}{\frac{1}{2(b_i - a_i)}} \leq 1, \quad i = 1, 2, \dots, n.$$

Now we want to construct a concentration ellipsoid  $(\beta - \beta_0)'T(\beta - \beta_0) = 1$  which encloses the cuboid and fulfils the following conditions:

- (i) The ellipsoid and the cuboid have the same center point,  $\beta_0 = \frac{1}{2}(a_1 + b_1, \dots, a_p + b_p)$ .
- (ii) The axes of the ellipsoid are parallel to the coordinate axes, that is,  $T = \text{diag}(t_1, \dots, t_p)$ .
- (iii) The corner points of the cuboid are on the surface of the ellipsoid, which means we have

$$\sum_{i=1}^p \left( \frac{a_i - b_i}{2} \right)^2 t_i = 1.$$

- (iv) The ellipsoid has minimal volume:

$$V = c_K \prod_{i=1}^p t_i^{\frac{1}{2}},$$

with  $c_p$  being a constant dependent on the dimension  $p$ .

We now include the linear restriction (iii) for the  $t_i$  by means of Lagrangian multipliers  $\lambda$  and solve (with

$$c_p^{-2} V_p^2 = \prod_{i=1}^p t_i^{-1})$$

$$\min_{\{t_i\}} \tilde{V} = \min_{t_i} \left\{ \prod_{i=1}^p t_i^{-1} - \lambda \left[ \sum_{i=1}^p \left( \frac{a_i - b_i}{2} \right)^2 t_i - 1 \right] \right\}.$$

The normal equations are then obtained as

$$\frac{\partial \tilde{V}}{\partial t_i} = -t_i^{-2} \prod_{i \neq j} t_i^{-1} - \lambda \left( \frac{a_j - b_j}{2} \right)^2 = 0$$

and 
$$\frac{\partial \tilde{V}}{\partial \lambda} = \sum \left( \frac{a_j - b_j}{2} \right)^2 t_i - 1 = 0.$$

From  $\frac{\partial \tilde{V}}{\partial t_i} = 0$ , we get

$$\begin{aligned} \lambda &= -t_i^{-2} \prod_{i \neq j} t_i^{-1} \left( \frac{2}{a_j - b_j} \right)^2 \quad (\text{for all } j = 1, 2, \dots, p) \\ &= -t_i^{-1} \prod_{i=1}^p t_i^{-1} \left( \frac{2}{a_j - b_j} \right)^2, \end{aligned}$$

and for any two  $i, j$  we obtain

$$t_i \left( \frac{a_j - b_j}{2} \right)^2 = t_j \left( \frac{a_j - b_j}{2} \right)^2,$$

and hence after summation according to  $\frac{\partial \tilde{V}}{\partial \lambda} = 0$  gives

$$\sum_{i=1}^p \left( \frac{a_j - b_j}{2} \right)^2 t_j = p t_j \left( \frac{a_j - b_j}{2} \right)^2 = 1.$$

This leads to the required diagonal elements of  $\mathbf{T}$ :

$$t_j = \frac{4}{p} (a_j - b_j)^{-2} \quad (j = 1, 2, \dots, p).$$

Hence, the optimal ellipsoid  $(\beta - \beta_0)' T (\beta - \beta_0) = 1$ , which contains the cuboid, has the center point vector

$$\beta'_0 = \frac{1}{2} (a_1 + b_1, \dots, a_p + b_p)$$

and the following matrix, which is positive definite for finite limits  $a_i, b_i (a_i \neq b_i)$ ,

$$T = \text{diag} \frac{4}{p} \left( (b_1 - a_1)^{-2}, \dots, (b_p - a_p)^{-2} \right).$$

**Interpretation:** The ellipsoid has a larger volume than the cuboid. Hence, the transition to an ellipsoid as a priori information represents a weakening but comes with easier mathematical handling.

**Example:** (Two real regressors) The center-point equation of the ellipsoid is (see Figure )

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

$$\text{or } (x, y) \begin{pmatrix} \frac{1}{a^2} & 0 \\ 0 & \frac{1}{b^2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 1$$

$$\text{with } T = \text{diag} \left( \frac{1}{a^2}, \frac{1}{b^2} \right) = \text{diag} (t_1, t_2)$$

$$\text{and the area } F = \pi ab = \pi t_1^{-\frac{1}{2}} t_2^{-\frac{1}{2}}.$$

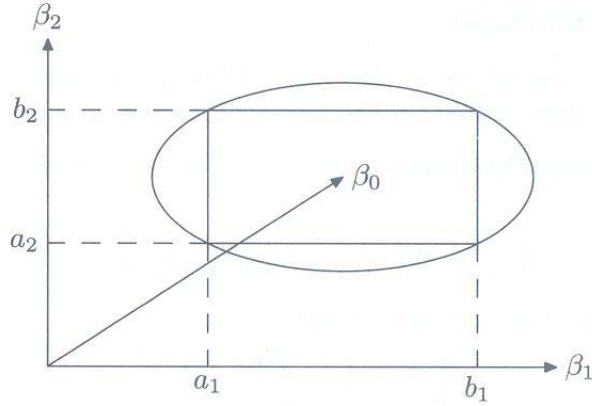


FIGURE: A priori rectangle and enclosing ellipsoid

### The Minimax Principle:

Consider the quadratic risk  $R(\hat{\beta}, \beta, A) = \text{tr} \left[ A E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' \right]$  and a class  $\{\hat{\beta}\}$  of estimators. Let

$B(\beta) \subset \mathbb{R}^p$  be a convex region of a priori restrictions for  $\beta$ . The criterion of the minimax estimator leads to the following.

**Definition:** An estimator  $b^* \in \{\hat{\beta}\}$  is called a minimax estimator of  $\beta$  if

An explicit solution can be achieved if the weight matrix is of the form  $A = aa'$  of rank 1.

Using the abbreviation,  $D_* = (S + k^{-1}\sigma^2 T)$  where  $S = X'X$ , we have the following result:

**Result:** In the model  $y = X\beta + \varepsilon, \varepsilon \sim N(0, \sigma^2 I)$ , with the restriction  $\beta'T\beta \leq k$  with  $T > 0$ , and the risk function  $R(\hat{\beta}, \beta, a)$ , the linear minimax estimator is of the following form:

$$\begin{aligned} b_* &= (X'X + k^{-1}\sigma^2 T)^{-1} X'y \\ &= D_*^{-1} X'y \end{aligned}$$

with the bias vector and covariance matrix as

$$\begin{aligned} \text{Bias}(b_*, \beta) &= -k^{-1}\sigma^2 D_*^{-1} T\beta, \\ V(b_*) &= \sigma^2 D_*^{-1} S D_*^{-1} \end{aligned}$$

and the minimax risk is

$$\sup_{\beta'T\beta \leq k} R(b_*, \beta, a) = \sigma^2 a' D_*^{-1} a.$$



**Result:** If the restrictions are  $(\beta - \beta_0)'T(\beta - \beta_0) \leq k$  with center point  $\beta_0 \neq 0$ , the linear minimax estimator is of the following form:

$$b_*(\beta_0) = \beta_0 + D_*^{-1}X'(y - X\beta_0)$$

with bias vector and covariance matrix as

$$\begin{aligned} \text{Bias}(b_*(\beta_0), \beta) &= -k^{-1}\sigma^2 D_*^{-1}T(\beta - \beta_0), \\ V(b_*(\beta_0)) &= V(b_*), \end{aligned}$$

and the minimax risk is

$$\sup_{(\beta - \beta_0)'T(\beta - \beta_0) \leq k} R(b_*(\beta_0), \beta, a) = \sigma^2 a' D_*^{-1} a.$$

**Interpretation:** A change of the center point of the a priori ellipsoid has an influence only on the estimator itself and its bias. The minimax estimator is not operational because  $\sigma^2$  is unknown. The smaller the value of  $k$ , the stricter is the a priori restriction for fixed  $T$ . Analogously, the larger the value of  $k$ , the smaller is the influence of  $\beta'T\beta \leq k$  on the minimax estimator. For the borderline case, we have

$$B(\beta) = \{\beta : \beta'T\beta \leq k\} \rightarrow \mathbb{R}^K \text{ as } k \rightarrow \infty$$

and  $\lim_{k \rightarrow \infty} b_* \rightarrow b = (X'X)^{-1}X'y.$

### Comparison of $b_*$ and $b$ :

**Minimax Risk:** Since the OLS estimator is unbiased, its minimax risk is

$$\sup_{\beta'T\beta \leq k} R(b, \cdot, a) = \sigma^2 a' S^{-1} a.$$

The linear minimax estimator  $b_*$  has a smaller minimax risk than the OLS estimator, and

$$\begin{aligned} R(b, \cdot, a) - \sup_{\beta'T\beta \leq k} R(b_*, \beta, a) \\ = \sigma^2 a' (S^{-1} - (k^{-1}\sigma^2 T + S)^{-1}) a \geq 0, \end{aligned}$$

since  $S^{-1} - (k^{-1}\sigma^2 T + S)^{-1} \geq 0$

Considering the superiority with MSE matrices, we get

$$\begin{aligned} M(b_*, \beta) &= V(b_*) + \text{Bias}(b_*, \beta) \text{Bias}(b_*, \beta)' \\ &= \sigma^2 D_*^{-1} (S + k^{-2}\sigma^2 T \beta \beta' T') D_*^{-1} \end{aligned}$$

Hence,  $b_*$  superior to  $b$  under the criterion of Loewner ordering when

$$\Delta(b, b_*) = V(b_x) - M(b_*, \beta) = \sigma^2 D_*^{-1} [D_* S^{-1} D_* - S - k^{-2} \sigma^2 T \beta \beta' T'] D_*^{-1} \geq 0,$$

which is possible if and only if

$$\begin{aligned} B &= D_* S^{-1} D_* - S - k^{-2} \sigma^2 T \beta \beta' T' \\ &= k^{-2} \sigma^4 T \left[ \left\{ S^{-1} + 2k \sigma^{-2} T^{-1} \right\} - \sigma^{-2} \beta \beta' \right] T \geq 0 \\ &= k^{-2} \sigma^4 T C^{\frac{1}{2}} \left[ I - \sigma^{-2} C^{-\frac{1}{2}} \beta \beta' C^{-\frac{1}{2}} \right] C^{-\frac{1}{2}} T \geq 0 \end{aligned}$$

with  $C = S^{-1} + 2k \sigma^{-2} T^{-1}$ . This is equivalent to

$$\sigma^{-2} \beta' (S^{-1} + 2k \sigma^{-2} T^{-1}) \geq 0.$$

Since  $(2k \sigma^{-2} T^{-1})^{-1} - (S^{-1} + 2k \sigma^{-2} T^{-1}) \geq 0$ ,

$$k^{-1} \leq \frac{2}{\beta' \beta}.$$

### Preliminary Test Estimation:

The statistical modeling of the data is usually done assuming that the model is correctly specified and the correct estimators are used for the purpose of estimation and drawing statistical inferences from a sample of data. Sometimes the prior information or constraints are available from outside the sample as non-sample information. The incorporation and use of such prior information along with the sample information lead to more efficient estimators provided it is correct. So the suitability of the estimator lies on the correctness of prior information. One possible statistical approach to check the correctness of prior information is through the framework of the test of hypothesis. For example, if prior information is available in the form of exact linear restrictions  $r = R\beta$ , there are two possibilities- either it is correct or incorrect. If the information is correct, then  $r = R\beta$  holds true in the model  $y = X\beta + \varepsilon$ , then the restricted regression estimator (RRE)

$\hat{\beta}_R = b + (X'X)^{-1} R' [R(X'X)^{-1} R']^{-1} (r - Rb)$  of is used, which is more efficient than OLSE

$b = (X'X)^{-1} X'y$  of  $\beta$ . Moreover, RRE satisfies the restrictions, i.e.  $R\hat{\beta}_R = r$ . On the other hand, when the information is incorrect, i.e.,  $r \neq R\beta$ , then OLSE is better than RRE. The truthfulness of prior information in terms of  $r = R\beta$  or  $r \neq R\beta$  is tested by the null hypothesis  $H_0 : R\beta = r$  using the  $F$ -statistics.

- If  $H_0$  is accepted at  $\alpha$  level of significance, then we conclude that  $R\beta = r$  and in such a situation, RRE is better than OLSE.
- On the other hand, if  $H_0$  is rejected at  $\alpha$  level of significance, then we conclude that  $R\beta \neq r$  and OLSE is better than RRE under such situations.

So when the exact content of the true sampling model is unknown, then the statistical model to be used is determined by a preliminary test of hypothesis using the available sample data. Such procedures are completed in two stages and are based on a test of hypothesis which provides a rule for choosing between the estimator based on the sample data, and the estimator is consistent with the hypothesis. This requires to make a test of the compatibility of OLSE (or maximum likelihood estimator) based on sample information only and RRE based on the linear hypothesis. The one can make a choice of estimator depending upon the outcome. Consequently, one can choose OLSE or RRE. Note that under the normality of random errors, the equivalent choice is made between the maximum likelihood estimator of  $\beta$  and the restricted maximum likelihood estimator of  $\beta$ , which has the same form as OLSE and RRE, respectively. So essentially a pre-test of hypothesis is done for  $H_0 : R\beta = r$  and based on that, a suitable estimator is chosen. This is called the pre-test procedure, which generates the pre-test estimator that, in turn, provides a rule to choose between restricted or unrestricted estimators.

One can also understand the philosophy behind the preliminary test estimation as follows. Consider the problem of an investigator who has a single data set and wants to estimate the parameters of a linear model that are known to lie in a high dimensional parametric space  $\Theta_1$ . However, the prior information about the parameter is available, and it suggests that the relationship may be characterized by a lower-dimensional parametric space  $\Theta_2 \subset \Theta_1$ . Under such uncertainty, if the parametric space  $\Theta_1$  is estimated by OLSE, the result from the over specified model will be unbiased but will have larger variance. Alternatively, the parametric space  $\Theta_2$  may incorrectly specify the statistical model and if estimated by OLSE will be biased. The bias may or may not outweigh the reduction in variance. If such uncertainty is represented in the form of the general linear hypothesis, this leads to pre-test estimators.

Let us consider the conventional pre-test estimator under the model  $y = X\beta + \varepsilon$  with usual assumptions and the general linear hypothesis  $H_0 : R\beta = r$  which can be tested by using  $F$  statistics. The null hypothesis  $H_0$  is rejected at  $\alpha$  level of significance when

$$u = F_{\text{calculated}} \geq F_{\alpha, p, n-p} = c$$

where the critical value  $c$  is determined for the given level of the test  $\alpha$  by

$$\int_c^\infty dF_{p, n-p} = P[F_{p, n-p} \geq c] = \alpha.$$

- If  $H_0$  is true, meaning thereby that the prior information is correct, then use RRE

$$\hat{\beta}_R = b + (X'X)^{-1} R' [R(X'X)^{-1} R']^{-1} (r - Rb) \text{ to estimate } \beta.$$

- If  $H_0$  is false, meaning thereby that the prior information is incorrect, then use OLSE

$$b = (X'X)^{-1} X'y \text{ to estimate } \beta.$$

Thus the estimator to be used depends on the preliminary test of significance and is of the form

$$\hat{\beta}_{PT} = \begin{cases} \hat{\beta}_R & \text{if } u < c \\ b & \text{if } u \geq c. \end{cases}$$

This estimator  $\hat{\beta}_{PT}$  called a preliminary test or pre-test estimator of  $\beta$ . Alternatively,

$$\begin{aligned} \hat{\beta}_{PT} &= \hat{\beta}_R \cdot I_{(0,c)}(u) + b \cdot I_{[c,\infty)}(u) \\ &= \hat{\beta}_R \cdot I_{(0,c)}(u) + [1 - I_{(0,c)}(u)] \cdot b \\ &= b - (b - \hat{\beta}_R) \cdot I_{(0,c)}(u) \\ &= b - (b - r) \cdot I_{(0,c)}(u) \end{aligned}$$

where the indicator functions are defined as

$$\begin{aligned} I_{(0,c)}(u) &= \begin{cases} 1 & \text{when } 0 < u < c \\ 0 & \text{otherwise} \end{cases} \\ I_{[c,\infty)}(u) &= \begin{cases} 1 & \text{when } u \geq c \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

- If  $\alpha = 0$ , then  $\hat{\beta}_{PT} = \hat{\beta}_R \cdot I_{[0,\infty)}(u) + b \cdot I_{(0)}(u) = \hat{\beta}_R$ .
- If  $\alpha = 1$  then  $\hat{\beta}_{PT} = \hat{\beta}_R \cdot I_{(0)}(u) + b \cdot I_{[0,\infty)}(u) = b$ .

Note that  $\alpha = 0$  and  $\alpha = 1$  indicate that the probability of type 1 error (i.e., rejecting  $H_0$  when it is true) is 0 and 1 respectively. So the entire area under the sampling distribution is the area of acceptance or the area of

rejection of null hypothesis. Thus the choice of  $\alpha$  has a crucial role to play in determining the sampling performance of the pre-test estimators. Therefore in a repeated sampling context, the data, the linear hypothesis, and the level of significance all determine the combination of the two estimators that are chosen on the average. The level of significance has an impact on the outcome of pretest estimator in the sense of determining the proportion of the time each estimator is used and in determining the sampling performance of pretest estimator.

We use the following result to derive the bias and risk of pretest estimator:

**Result 1:** If the  $K \times 1$  random vector,  $Z / \sigma$ , is distributed as a multivariate normal random vector with mean  $\delta / \sigma$  and covariance matrix  $I_k$  and is independent of  $\chi^2_{(n-k)}$  then

$$E \left[ I_{(0,c)} \left( \frac{Z'Z}{\sigma^2 \chi^2_{(n-K)}} \frac{n-K}{K} \right) \frac{Z}{\sigma} \right] = \left( \frac{\delta}{\sigma} \right) P \left[ \frac{\chi^2_{(K+2,\lambda)}}{\chi^2_{(n-K)}} \leq c^* \right],$$

where  $\lambda = \delta' \delta / 2\sigma^2$  and  $c^* = cK / (n-K)$ .

**Result 2:** If the  $K \times 1$  random vector,  $Z / \sigma$  is distributed as a multivariate normal random vector with mean  $\delta / \sigma$  and covariance  $I_k$  and is independent of  $\chi^2_{(n-k)}$  then

$$\begin{aligned} E \left[ I_{(0,c^*)} \left( \frac{Z'Z}{\sigma^2 \chi^2_{(n-K)}} \right) \left( \frac{Z}{\sigma} \right) \left( \frac{Z}{\sigma} \right)' \right] &= KE \left[ I_{(0,c^*)} \left( \frac{\chi^2_{(K+2,\delta'\delta/2\sigma^2)}}{\chi^2_{(n-K)}} \right) \right] + \frac{\delta'\delta}{\sigma^2} E \left[ I_{(0,c^*)} \frac{\chi^2_{(K+4,\delta'\delta/2\sigma^2)}}{\chi^2_{(n-K)}} \right] \\ &= KP \left[ \frac{\chi^2_{(K+2,\delta'\delta/2\sigma^2)}}{\chi^2_{(n-K)}} \leq \frac{cK}{n-K} \right] + \frac{\delta'\delta}{\sigma^2} P \left[ \frac{\chi^2_{(K+4,\delta'\delta/2\sigma^2)}}{\chi^2_{(n-K)}} \leq \frac{cK}{n-K} \right], \end{aligned}$$

where  $c^* = cK / (n-K)$ .

Using these results, we can find the bias and risk of  $\hat{\beta}_{PT}$  follows:

**Bias:**

$$\begin{aligned} E(\hat{\beta}_{PT}) &= E(b) - E \left[ I_{(0,c)}(u) \cdot (b-r) \right] \\ &= \beta - \delta P \left[ \frac{\chi^2_{(p+2,\delta'\delta/2\sigma^2)}}{\chi^2_{(n-p,\delta'\delta/2\sigma^2)}} \leq c \frac{p}{n-p} \right] \\ &= \beta - \delta P \left[ F_{(p+2, n-p, \delta'\delta/2\sigma^2)} \leq c \right] \end{aligned}$$

where  $\delta = r - R\beta$ ,  $\chi^2_{(p+2, \delta'\delta/2\sigma^2)}$  denotes the non-central  $\chi^2$  distribution with noncentrality parameter  $\delta'\delta/2\sigma^2$ ,  $F_{(p+2, n-p, \delta'\delta/2\sigma^2)}$  denotes the non-central  $F$  distribution with noncentrality parameter  $\delta'\delta/2\sigma^2$ .

Thus, if  $\delta = 0$ , the pretest estimator is unbiased. Note that the size of bias is affected by the probability of a random variable with a non-central  $F$ - distribution being less than a constant that is determined by the level of the test, the number of hypothesis and the degree of hypothesis error  $\delta$ . Since the probability is always less than or equal to one, so  $bias(\hat{\beta}_{PT}) \leq bias(b)$ .

### Risk:

The risk of pretest estimator is obtained as

$$\begin{aligned} \rho(\beta, \hat{\beta}_{PT}) &= E \left[ (\hat{\beta}_{PT} - \beta)' (\hat{\beta}_{PT} - \beta) \right] \\ &= E \left[ (b - \beta - I_{(0,c)}(u)(b - r))' (b - \beta - I_{(0,c)}(u)(b - r)) \right] \\ &= E \left[ (b - \beta)' (b - \beta) \right] - E \left[ I_{(0,c)}(u)(b - \beta)' (b - \beta) \right] + E \left[ I_{(0,c)}(u) \right] \delta' \delta \\ &= \sigma^2 p + (2\delta' \delta - \sigma^2 p) P \left[ \frac{\chi^2_{(p+2, \delta'\delta/2\sigma^2)}}{\chi^2_{(n-p)}} \leq \frac{cp}{n-p} \right] - \delta' \delta P \left[ \frac{\chi^2_{(p+4, \delta'\delta/2\sigma^2)}}{\chi^2_{(n-p)}} \leq \frac{cp}{n-p} \right] \end{aligned}$$

or compactly,

$$\rho(\beta, \hat{\beta}_{PT}) = \sigma^2 p + (2\delta' \delta - \sigma^2 p) l(2) - \delta' \delta l(4)$$

where

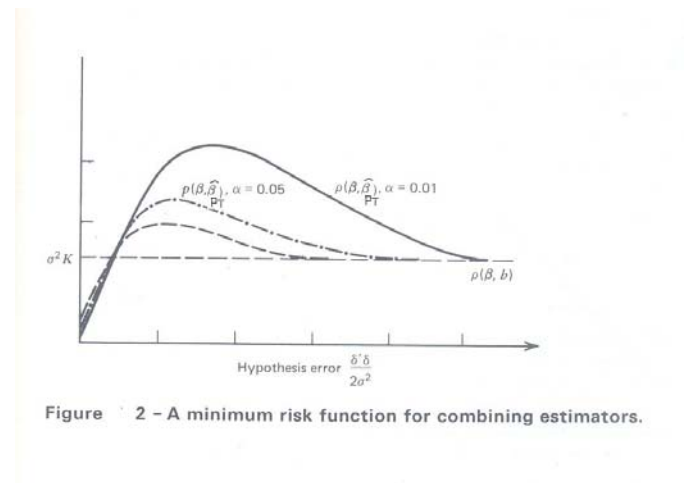
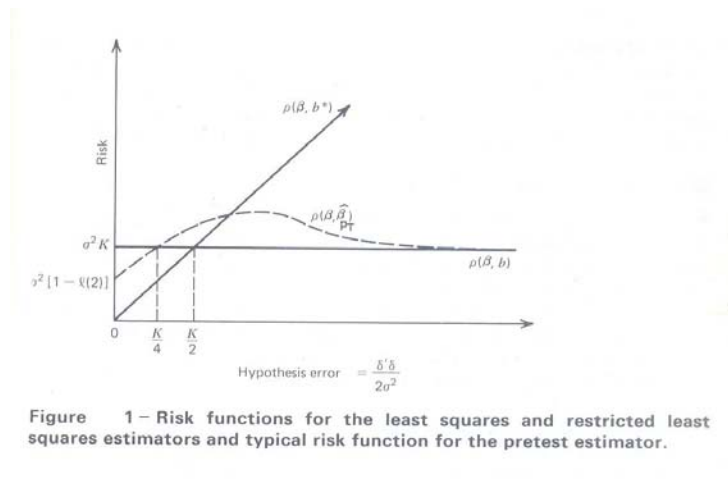
$$l(2) = \frac{\chi^2_{(p+2, \delta'\delta/2\sigma^2)}}{\chi^2_{(n-p)}}, \quad l(4) = \frac{\chi^2_{(p+4, \delta'\delta/2\sigma^2)}}{\chi^2_{(n-p)}}, \quad 0 < l(4) < l(2) < 1.$$

The risk function implies the following results:

1. If the restrictions are correct and  $\delta = 0$ , the risk of the pretest estimator  $\sigma^2 K[1-l(2)]$  where  $0 < [1-l(2)] < 1$  for  $0 < c < \infty$ . Therefore, the pretest estimator has risk less than that of the least-squares estimator at the origin  $\delta = 0$ , and the decrease in risk depends on the level of significance  $\alpha$  and, correspondingly, on the critical value of the test  $c$ .
2. As the hypothesis error  $\beta - r = \delta$ , and thus  $\delta'\delta/2\sigma^2$ , increases and approaches infinity,  $l(\cdot)$  and  $\delta\delta'l(\cdot)$  approach zero. The risk of the pretest estimator, therefore, approaches  $\sigma^2 p$ , the risk of the unrestricted least squares estimator.

3. As the hypothesis error grows, the risk of the pretest estimator increases obtains a maximum after crossing the risk of the least-squares estimator, and then monotonically decreases to approach  $\sigma^2 p$  the risk of the OLSE.
4. The pretest estimator risk function defined on the  $\delta'\delta/2\sigma^2$  parameter spaces crosses the risk function of the least-squares estimator within the bounds  $p/4 \leq \delta'\delta/2\sigma^2 \leq p/2$ .

The sampling characteristic of the preliminary test estimator are summarized in Figure 1.



From these results, we see that the pretest estimator does well relative to OLSE if the hypothesis is correctly specified. However, in the  $\delta'\delta/2\sigma^2$  space representing the range of hypothesis are correctly specified. However, in the  $\delta'\delta/2\sigma^2$  space representing the range of hypothesis errors, the pretest estimator is inferior to the least-squares estimator over an infinite range of the parameter space. In figures 1 and 2, there is a range of the parameter space which the pretest estimator has the risk that is inferior to (greater than) that of both the unrestricted and restricted least squares estimators. No one estimator depicted in Figure 1 dominates the other competitors. In addition, in applied problems, the hypothesis errors, and thus the correct  $\beta$  in the specification error parameter space, are seldom known. Consequently, the choice of estimator is unresolved.

### The Optimal Level of Significance

The form of the pretest estimator involves, for evaluation purposes, the probabilities of ratios of random variables  $l(2)$  and  $l(4)$  being less than a constant that depends on the critical value of the test  $c$  or on the level of statistical significance  $\alpha$ . Thus as  $\alpha \rightarrow 0$  the probabilities  $l(\cdot) \rightarrow 1$ , and the risk of the pretest estimator approaches that of the restricted regression estimator  $\hat{\beta}_R$ . In contrast, as  $\alpha \rightarrow 1$ ,  $l(\cdot)$  approaches zero and the risk of the pretest estimator approaches that of the least-squares estimator  $b$ . The choice of  $\alpha$ , which has a crucial impact on the performance of the pretest estimator, is portrayed in Figure 3.

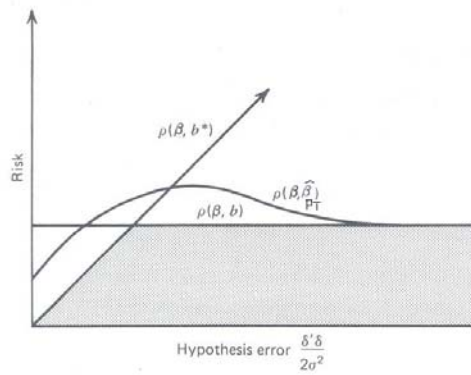


Figure 3 – Impact of choice of the critical value of the test on the value of the Pretest risk function.

Since the investigator is usually unsure of the degree of hypothesis specification error, and thus is unsure of the appropriate point in the  $\delta$  space for evaluating the risk, the best of worlds would be to have a rule that mixes the unrestricted and restricted estimators so as to minimize risk regardless of the relevant specification error  $\delta'\delta/2\sigma^2$ . Thus the risk function traced out by the cross-hatched area in Figure 2 is relevant. Unfortunately, the risk of the pretest estimator, regardless of the choice of  $\alpha$ , is always equal to or greater than the minimum risk function for some range of the parameter space. Given this result, one criterion that has been proposed for choosing the  $\alpha$  level might be to select the critical value  $c$  that would minimize the maximum regret of not being on the minimum risk function, reflected by the boundary of the shaded area. Another criterion that has been proposed for choosing  $\alpha$  is to minimize the average regret over the whole  $\delta'\delta/2\sigma^2$  space. Each of these criteria leads to different conclusions or rules for choice, and the question concerning the optimal level of the test is still open. One obvious thing is that conventional choices of 0.05 and 0.01 may have rather severe statistical consequences.