

Restricted Least Squares, Hypothesis Testing, and Prediction in the Classical Linear Regression Model

A. Introduction and assumptions

The classical linear regression model can be written as

$$y = X\beta + \epsilon \quad (1)$$

or

$$y_t = x_t' \beta + \epsilon_t \quad (t=1, \dots, n). \quad (2)$$

where x_t' is the t th row of the matrix X or simply as

$$y_t = x_t \beta + \epsilon_t \quad (t=1, \dots, n). \quad (3)$$

where it is implicit that x_t is a row vector containing the regressors for the t th time period. The classical assumptions on the model can be summarized as

$$\begin{aligned} I \quad & y = X\beta + \epsilon \\ II \quad & E(\epsilon | X) = 0 \\ III \quad & E(\epsilon\epsilon' | X) = \sigma^2 I \\ IV \quad & X \text{ is a nonstochastic matrix of rank } k \\ V \quad & \epsilon \sim N(0; \Sigma = \sigma^2 I) \end{aligned} \quad (4)$$

Assumption V as written implies II and III. With normally distributed disturbances, the joint density (and therefore likelihood function) of y is

$$\begin{aligned} L(y; \mu, \Sigma = \sigma^2 I) &= \frac{e^{-\frac{1}{2}(y-X\beta)' \Sigma^{-1}(y-X\beta)}}{(2\pi)^{n/2} |\Sigma|^{\frac{1}{2}}} \\ &= \frac{e^{-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)}}{(2\pi)^{n/2} |\sigma^2 I|^{\frac{1}{2}}} \\ &= \frac{e^{-\frac{1}{2\sigma^2}(y-X\beta)'(y-X\beta)}}{(2\pi)^{n/2} (\sigma^2)^{\frac{n}{2}}} \end{aligned} \quad (5)$$

The natural log of the likelihood function is given by

$$\begin{aligned}
\ell = \ln L &= - \frac{(y - X\beta)'(y - X\beta)}{2\sigma^2} - \frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 \\
&= - \frac{y'y - 2\beta'X'y + \beta'X'X\beta}{2\sigma^2} - \frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2
\end{aligned} \tag{6}$$

Maximum likelihood estimators are obtained by setting the derivatives of (6) equal to zero and solving the resulting $k+1$ equations for the k β 's and σ^2 . These first order conditions for the M.L estimators are

$$\frac{\partial \ell}{\partial \beta'} = -\frac{1}{2\sigma^2} (-2X'y + 2(X'X)\beta) = 0 \tag{7}$$

$$\frac{\partial \ell}{\partial \sigma^2} = \frac{1}{2} \left(\frac{1}{\sigma^2} \right)^2 (y - X\beta)'(y - X\beta) - \frac{n}{2} \frac{1}{\sigma^2} = 0$$

Solving we obtain

$$\begin{aligned}
\beta &= (X'X)^{-1}X'y \\
\hat{\sigma}^2 &= \frac{1}{n} (y - X\beta)'(y - X\beta) \\
&= \frac{e'e}{n} = \frac{\sum e_i^2}{n}
\end{aligned} \tag{8}$$

The ordinary least squares estimator is obtained by minimizing the sum of squared errors which is defined by

$$\begin{aligned}
SSE(\beta) &= \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - x_i'\beta)^2 \\
&= [e_1 \ e_2 \ \dots \ e_n] \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \\
&= e'e \\
&= (y - X\beta)'(y - X\beta) \\
&= y'y - \beta'X'y - y'X\beta + \beta'X'X\beta \\
&= y'y - 2\beta'X'y + \beta'X'X\beta
\end{aligned} \tag{9}$$

The necessary condition for $SSE(\beta)$ to be a minimum is that

$$\begin{aligned}\frac{\partial SSE(\hat{\beta})}{\partial \hat{\beta}'} &= 0 \\ \Rightarrow -2X'y + 2X'X\hat{\beta} &= 0\end{aligned}\tag{10}$$

This gives the normal equations which can then be solved to obtain the least squares estimator

$$\begin{aligned}X'X\hat{\beta} &= X'y \\ \Rightarrow \hat{\beta} &= (X'X)^{-1}X'y\end{aligned}\tag{11}$$

The maximum likelihood estimator of β is the same as the least squares estimator.

B. Restricted least squares

1. Linear restrictions on β

Consider a set of m linear constraints on the coefficients denoted by

$$R\beta = r \quad R \text{ is } m \times k, \quad r \text{ is } m \times 1\tag{12}$$

Restricted least squares estimation or restricted maximum likelihood estimation consists of minimizing the objective function in (9) or maximizing the objective function in (6) subject to the constraint in (12).

2. Constrained maximum likelihood estimates

Given that there is no constraint on σ^2 , we can differentiate equation 6 with respect to σ^2 to get an estimator of σ^2 as a function of the restricted estimator of β . Doing so we obtain'

$$\begin{aligned}\frac{\partial \ell}{\partial \sigma^2} &= -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2} \left(\frac{1}{\hat{\sigma}^2} \right)^2 (y - X\beta^c)' (y - X\beta^c) = 0 \\ \Rightarrow \hat{\sigma}^2 &= \frac{1}{n} (y - X\beta^c)' (y - X\beta^c)\end{aligned}\tag{13}$$

where β^c is the constrained maximum likelihood estimator. Now substitute this estimator for σ^2 back into the log likelihood equation (6) and simplify to obtain

$$\begin{aligned}\ell = \ln L &= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{(y - X\beta)' (y - X\beta)}{2\sigma^2} \\ &= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \left[\frac{1}{n} (y - X\beta^c)' (y - X\beta^c) \right] - \frac{(y - X\beta^c)' (y - X\beta^c)}{2 \left[\frac{1}{n} (y - X\beta^c)' (y - X\beta^c) \right]} \\ &= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \left[\frac{1}{n} (y - X\beta^c)' (y - X\beta^c) \right] - \frac{n}{2}\end{aligned}\tag{14}$$

Note that the concentrated likelihood function (as opposed to the concentrated log likelihood function) is given by

$$\begin{aligned}\ln L &= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \left[\frac{1}{n} (y - X\beta^c)' (y - X\beta^c) \right] - \frac{n}{2} \\ \Rightarrow L &= (2\pi)^{-\frac{n}{2}} \left[\frac{1}{n} (y - X\beta^c)' (y - X\beta^c) \right]^{-\frac{n}{2}} e^{-\frac{n}{2}}\end{aligned}\quad (15)$$

The maximization problem defining the restricted estimator can then be stated as

$$\begin{aligned}\max_{\beta^c} \ell &= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \left[\frac{1}{n} (y - X\beta^c)' (y - X\beta^c) \right] - \frac{n}{2} \\ \text{s.t. } R\beta^c &= r\end{aligned}\quad (16)$$

Clearly we maximize this likelihood function by minimizing the sum of squared errors $(y - X\beta^c)'(y - X\beta^c)$. To maximize this subject to the constraint we form the Lagrangian function where λ' is an $m \times 1$ vector of Lagrangian multipliers

$$\mathcal{L} = y'y - 2\beta^{c'}X'y + \beta^{c'}X'X\beta^c - \lambda'(r - R\beta^c) \quad (17)$$

Differentiation with respect to $\beta^{c'}$ and λ yields the conditions

$$\begin{aligned}-2X'y + 2X'X\beta^c + R'\lambda &= 0 \\ R\beta^c - r &= 0\end{aligned}\quad (18)$$

Now multiply the first equation in (18) by $R(X'X)^{-1}$ to obtain

$$-2R(X'X)^{-1}X'y + 2R\beta^c + R(X'X)^{-1}R'\lambda = 0 \quad (19)$$

Now solve this equation for λ substituting $\hat{\beta} = (X'X)^{-1}X'y$ as appropriate

$$\begin{aligned}-2R(X'X)^{-1}X'y + 2R\beta^c + R(X'X)^{-1}R'\lambda &= 0 \\ \Rightarrow R(X'X)^{-1}R'\lambda &= 2R(X'X)^{-1}X'y - 2R\beta^c \\ &= 2R\hat{\beta} - 2R\beta^c \\ \Rightarrow \lambda &= \left(R(X'X)^{-1}R' \right)^{-1} (2R\hat{\beta} - 2R\beta^c) \\ &= -2 \left(R(X'X)^{-1}R' \right)^{-1} (r - R\hat{\beta})\end{aligned}\quad (20)$$

The last step follows because $R\beta^c = r$. Now substitute this back into the first equation in (18) to obtain

$$\begin{aligned}
& -2X'y + 2X'X\beta^c + R'\lambda = 0 \\
& \Rightarrow -2X'y + 2X'X\beta^c + R'[-2(R(X'X)^{-1}R')^{-1}(r - R\beta)] = 0 \\
& \Rightarrow X'X\beta^c = X'y + R'(R(X'X)^{-1}R')^{-1}(r - R\beta) \\
& \Rightarrow \beta^c = (X'X)^{-1}X'y + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\beta) \\
& \quad = \beta + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\beta)
\end{aligned} \tag{21}$$

With normally distributed errors in the model, the maximum likelihood and least squares estimates of the constrained model are the same.

We can rearrange (21) in the following useful fashion

$$\begin{aligned}
\beta^c &= \beta + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\beta) \\
\Rightarrow \beta^c - \beta &= (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\beta) \\
\Rightarrow (X'X)(\beta^c - \beta) &= R'(R(X'X)^{-1}R')^{-1}(r - R\beta)
\end{aligned} \tag{22}$$

Now multiply both sides of (22) by $(\beta^c - \beta)' = (r - R\beta)'(R(X'X)^{-1}R')^{-1}R(X'X)^{-1}$ to obtain

$$\begin{aligned}
(\beta^c - \beta)'(X'X)(\beta^c - \beta) &= (r - R\beta)'(R(X'X)^{-1}R')^{-1}R(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\beta) \\
&= (r - R\beta)'(R(X'X)^{-1}R')^{-1}(r - R\beta)
\end{aligned} \tag{23}$$

We can rearrange equation 21 in another useful fashion by multiplying both sides by X and then subtracting both sides from y. Doing so we obtain

$$\begin{aligned}
X\beta^c &= X\beta + X(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\beta) \\
\Rightarrow y - X\beta^c &= y - X\beta - X(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\beta) \\
\Rightarrow u &= e - X(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\beta)
\end{aligned} \tag{24}$$

where u is the estimated residual from the constrained regression. Consider also $u'u$ which is the sum of squared errors from the constrained regression.

$$\begin{aligned}
u'u &= (e' - (r - R\beta)'(R(X'X)^{-1}R')^{-1}R(X'X)^{-1}X')(e - X(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\beta)) \\
&= e'e - e'X(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\beta) - (r - R\beta)'(R(X'X)^{-1}R')^{-1}R(X'X)^{-1}X'e \\
&\quad + (r - R\beta)'(R(X'X)^{-1}R')^{-1}R(X'X)^{-1}X'X(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\beta) \\
&= e'e - e'X(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\beta) - (r - R\beta)'(R(X'X)^{-1}R')^{-1}R(X'X)^{-1}X'e \\
&\quad + (r - R\beta)'(R(X'X)^{-1}R')^{-1}(r - R\beta)
\end{aligned} \tag{25}$$

where e is the estimated residual vector from the unconstrained model. Now remember that in ordinary least squares $X'e = 0$ as can be seen by rewriting equation 10 as follows

$$\begin{aligned}
2X'y - 2X'X\hat{\beta} &= 0 \\
\Rightarrow X'(y - X\hat{\beta}) &= X'e = 0
\end{aligned} \tag{26}$$

Using this information in equation 25 we obtain

$$\begin{aligned}
u'u &= e'e - e'X(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\hat{\beta}) - (r - R\hat{\beta})'(R(X'X)^{-1}R')^{-1}R(X'X)^{-1}X'e \\
&\quad + (r - R\hat{\beta})'(R(X'X)^{-1}R')^{-1}(r - R\hat{\beta}) \\
&= e'e + (r - R\hat{\beta})'(R(X'X)^{-1}R')^{-1}(r - R\hat{\beta}) \\
\Rightarrow u'u - e'e &= (r - R\hat{\beta})'(R(X'X)^{-1}R')^{-1}(r - R\hat{\beta})
\end{aligned} \tag{27}$$

Thus the difference in the sum of squared errors in the constrained and unconstrained models can be written as a quadratic form in the difference between $R\hat{\beta}$ and r where $\hat{\beta} = (X'X)^{-1}X'y$ is the unconstrained ordinary least squares estimate.

Equation 21 can be rearranged in yet another fashion that will be useful in finding the variance of the constrained estimator. First write the ordinary least square estimator as a function of β and ϵ as follows

$$\begin{aligned}
\hat{\beta} &= (X'X)^{-1}X'y \\
&= (X'X)^{-1}X'(X\beta + \epsilon) \\
&= \beta + (X'X)^{-1}X'\epsilon
\end{aligned} \tag{28}$$

Then substitute this expression for $\hat{\beta}$ in equation 21 as follows

$$\begin{aligned}
\beta^c &= \hat{\beta} + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\hat{\beta}) \\
&= \beta + (X'X)^{-1}X'\epsilon + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R(\beta + (X'X)^{-1}X'\epsilon)) \\
&= \beta + (X'X)^{-1}X'\epsilon + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\beta - R(X'X)^{-1}X'\epsilon) \\
\Rightarrow \beta^c - \beta &= (X'X)^{-1}X'\epsilon + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(-R(X'X)^{-1}X'\epsilon), \quad r = R\beta \\
\Rightarrow \beta^c - \beta &= (X'X)^{-1}X'\epsilon - (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}R(X'X)^{-1}X'\epsilon
\end{aligned} \tag{29}$$

Now define the matrix M^c as follows

$$M^c = \left[I - (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}R \right] \tag{30}$$

We can then write $\beta^c - \beta$ as

$$\begin{aligned}
\beta^c - \beta &= (X'X)^{-1}X'\epsilon - (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}R(X'X)^{-1}X'\epsilon \\
&= \left[I - (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}R \right] (X'X)^{-1}X'\epsilon \\
&= M^c (X'X)^{-1}X'\epsilon \\
&= M^c [\hat{\beta} - \beta]
\end{aligned} \tag{31}$$

3. Statistical properties of the restricted least squares estimates

a. expected value of β^c

$$\begin{aligned}
\beta^c &= \hat{\beta} + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\hat{\beta}) \\
E(\beta^c) &= E(\hat{\beta}) + E\left[(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\hat{\beta})\right] \\
&= \beta + \left[(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - RE(\hat{\beta}))\right] \\
&= \beta + \left[(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(r - R\beta)\right] \\
&= \beta, \text{ if the constraint is true}
\end{aligned} \tag{32}$$

b. variance of β^c

$$\begin{aligned}
Var(\beta^c) &= E[(\beta^c - \beta)(\beta^c - \beta)'] \\
E(\beta^c - \beta)(\beta^c - \beta)' &= E[M^c(X'X)^{-1}X'\epsilon\epsilon'X(X'X)^{-1}M^{c'}] \\
&= M^c(X'X)^{-1}X'E(\epsilon\epsilon')X(X'X)^{-1}M^{c'} \\
&= M^c(X'X)^{-1}X'\sigma^2IX(X'X)^{-1}M^{c'} \\
&= \sigma^2 M^c(X'X)^{-1}X'X(X'X)^{-1}M^{c'} \\
&= \sigma^2 M^c(X'X)^{-1}M^{c'}
\end{aligned} \tag{33}$$

The matrix M^c is not symmetric, but it is idempotent as can be seen by multiplying it by itself.

$$\begin{aligned}
M^c M^c &= \left[I - (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}R \right] \left[I - (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}R \right] \\
&= I - (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}R - (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}R \\
&\quad + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}R(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}R \\
&= I - 2(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}R + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}R \\
&= I - (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}R
\end{aligned} \tag{34}$$

Now consider the expression for the variance of β^c . We can write it out and simplify to obtain

$$\begin{aligned}
Var(\beta^r) &= \sigma^2 M' (X'X)^{-1} M'' \\
&= \sigma^2 \left[I - (X'X)^{-1} R' (R(X'X)^{-1} R')^{-1} R \right] (X'X)^{-1} \left[I - R' (R(X'X)^{-1} R')^{-1} R (X'X)^{-1} \right] \\
&= \sigma^2 \left[(X'X)^{-1} \right] \\
&\quad - \sigma^2 \left[(X'X)^{-1} R' (R(X'X)^{-1} R')^{-1} R (X'X)^{-1} \right] \\
&\quad - \sigma^2 \left[(X'X)^{-1} R' (R(X'X)^{-1} R')^{-1} R (X'X)^{-1} \right] \\
&\quad + \sigma^2 \left[(X'X)^{-1} R' (R(X'X)^{-1} R')^{-1} R (X'X)^{-1} R' (R(X'X)^{-1} R')^{-1} R (X'X)^{-1} \right] \\
&= \sigma^2 \left[(X'X)^{-1} - 2 (X'X)^{-1} R' (R(X'X)^{-1} R')^{-1} R (X'X)^{-1} + (X'X)^{-1} R' (R(X'X)^{-1} R')^{-1} R (X'X)^{-1} \right] \\
&= \sigma^2 \left[(X'X)^{-1} - (X'X)^{-1} R' (R(X'X)^{-1} R')^{-1} R (X'X)^{-1} \right] \\
&= \sigma^2 \left[I - R' (R(X'X)^{-1} R')^{-1} R (X'X)^{-1} \right] (X'X)^{-1} \\
&= \sigma^2 M' (X'X)^{-1}
\end{aligned} \tag{35}$$

We can also write this is another useful form

$$\begin{aligned}
Var(\beta^r) &= \sigma^2 \left[(X'X)^{-1} - (X'X)^{-1} R' (R(X'X)^{-1} R')^{-1} R (X'X)^{-1} \right] \\
&= \sigma^2 (X'X)^{-1} - \sigma^2 (X'X)^{-1} R' (R(X'X)^{-1} R')^{-1} R (X'X)^{-1}
\end{aligned} \tag{36}$$

The variance of the restricted least squares estimator is thus the variance of the ordinary least squares estimator minus a positive semi-definite matrix, implying that the restricted least squares estimator has a lower variance than the OLS estimator.

4. Testing the restrictions on the model using estimated residuals

We showed previously (equation 109 in the section on statistical inference) that

$$\frac{\frac{(r - R\beta)' [R(X'X)^{-1} R']^{-1} (r - R\beta)}{m}}{\frac{e'e}{n - k}} \sim F(m, n - k) \tag{37}$$

Consider the numerator in equation 37. It can be written in terms of the residuals from the restricted and unrestricted models using equation 27

$$\begin{aligned}
u'u &= e'e + (r - R\beta)' (R(X'X)^{-1} R')^{-1} (r - R\beta) \\
\Rightarrow u'u - e'e &= (r - R\beta)' (R(X'X)^{-1} R')^{-1} (r - R\beta) \\
\frac{u'u - e'e}{m} &= \frac{(r - R\beta)' (R(X'X)^{-1} R')^{-1} (r - R\beta)}{m} \\
\Rightarrow \frac{\frac{u'u - e'e}{m}}{\frac{e'e}{n - k}} &\sim F(m, n - k)
\end{aligned} \tag{38}$$

Denoting the sum of squared residuals from a particular model by $SSE(\beta)$ we obtain

$$\left(\frac{\frac{SSE(\beta^c) - SSE(\beta)}{m}}{\frac{SSE(\beta)}{n - k}} \right) \sim F(m, n - k) \quad (39)$$

Rather than performing the hypothesis test by inverting the matrix $[R(X'X)^{-1}R']$ and then pre and post multiplying by $r - R\beta$, we simply run two different regressions and compute the F statistic from the constrained and unconstrained residuals.

The form of the test statistic in equation 39 is referred to as a test based on the change in the objective function. A number of tests fall in this category. The idea is to compare the sum of squares in the constrained and unconstrained models. If the restriction causes SSE to be significantly larger than otherwise, this is evidence that the data do not satisfy the restriction and we reject the hypothesis that the restriction holds. The general procedure for such tests is to run two regressions as follows.

- 1) Estimate the regression model without imposing any constraints on the vector β . Let the associated sum of squared errors (SSE) and degrees of freedom be denoted by SSE and $(n - k)$, respectively.
- 2) Estimate the same regression model where the β is constrained as specified by the hypothesis. Let the associated sum of squared errors (SSE) and degrees of freedom be denoted by $SSE(\beta^c)$ and $(n - k)^c$, respectively.
- 3) Perform test using the following statistic

$$\left(\frac{\frac{SSE(\beta^c) - SSE(\beta)}{m}}{\frac{SSE(\beta)}{n - k}} \right) \sim F(m, n - k) \quad (40)$$

where $m = (n - k)^c - (n - k)$ is the number of independent restrictions imposed on β by the hypothesis. For example, if the hypothesis was $H_0: \beta_2 + \beta_3 = 4, \beta_4 = 0$, then the numerator degrees of freedom (q) is equal to 2. If the hypothesis is valid, then $SSE(\beta^c)$ and (SSE) should not be significantly different from each other. Thus, we reject the constraint if the F value is large. Two useful references on this type of test are Chow and Fisher.

5. Testing the restrictions on the model using a likelihood ratio (LR) test

a. idea and definition

The likelihood ratio (LR) test is a common method of statistical inference in classical statistics. The LR test statistic reflects the compatibility between a sample of data and the null hypothesis through a comparison of the constrained and unconstrained likelihood functions. It is based on determining whether there has been a significant reduction in the value of the likelihood (or log-likelihood) value as a result of imposing the hypothesized constraints on the parameters θ in the estimation process. Formally let the random sample (X_1, X_2, \dots, X_n) have the joint probability density function $f(x_1, x_2, \dots, x_n; \theta)$ and the associated likelihood function $L(\theta; x_1, x_2, \dots, x_n)$. The generalized likelihood ratio (GLR) is defined as

$$\lambda(x_1, x_2, \dots, x_n) = \left(\frac{\sup_{\theta \in H_0} L(\theta; x_1, x_2, \dots, x_n)}{\sup_{\theta \in H_0 \cup H_a} L(\theta; x_1, x_2, \dots, x_n)} \right) \quad (41)$$

where sup denotes the supremum of the function over the set of parameters satisfying the null hypothesis (H_0), or the set of parameters that would satisfy either the null or alternative (H_a) hypothesis. A generalized likelihood ratio test for testing H_0 against the alternative H_a is given by the following test rule:

$$\text{Reject } H_0 \text{ if and only if } \lambda(x_1, x_2, \dots, x_n) \leq c \quad (42)$$

where c is the critical value from a yet to be determined distribution. For an α test the constant c is chosen to satisfy

$$\sup_{\theta \in H_0} \pi(\theta) = \sup_{\theta \in H_0} P(\lambda(x_1, x_2, \dots, x_n) \leq c; \theta) = \alpha \quad (43)$$

where $\pi(\theta)$ is the power function of the statistical test.

We use the term “generalized” likelihood ratio as compared to likelihood ratio to indicate that the two likelihood functions in the ratio are “optimized” with respect to θ over the two different domains. The literature often refers to this as a likelihood ratio test without the modifier generalized and we will often follow that convention.

b. likelihood ratio test for the classical normal linear regression model

Consider the null hypothesis in the classical normal linear regression model $R\beta = r$. The likelihood function evaluated at the restricted least squares estimates from equation 15 is

$$L^{c*} = (2\pi)^{-\frac{n}{2}} \left[\frac{1}{n} (y - X\beta^c)' (y - X\beta^c) \right]^{-\frac{n}{2}} e^{-\frac{n}{2}} \quad (44)$$

In an analogous manner we can write the likelihood function evaluated at the OLS estimates as

$$L^* = (2\pi)^{-\frac{n}{2}} \left[\frac{1}{n} (y - X\hat{\beta})' (y - X\hat{\beta}) \right]^{-\frac{n}{2}} e^{-\frac{n}{2}} \quad (45)$$

The generalized likelihood ratio statistic is then

$$\begin{aligned}
\lambda(x_1, x_2, \dots, x_n) &= \frac{\left((2\pi)^{-\frac{n}{2}} \left[\frac{1}{n} (y - X\beta^c)' (y - X\beta^c) \right]^{-\frac{n}{2}} e^{-\frac{n}{2}} \right)}{\left((2\pi)^{-\frac{n}{2}} \left[\frac{1}{n} (y - X\hat{\beta})' (y - X\hat{\beta}) \right]^{-\frac{n}{2}} e^{-\frac{n}{2}} \right)} \\
&= \frac{[(y - X\beta^c)' (y - X\beta^c)]^{-\frac{n}{2}}}{[(y - X\hat{\beta})' (y - X\hat{\beta})]^{-\frac{n}{2}}} \\
&= \left(\frac{(y - X\beta^c)' (y - X\beta^c)}{(y - X\hat{\beta})' (y - X\hat{\beta})} \right)^{-\frac{n}{2}}
\end{aligned} \tag{46}$$

We reject the null hypothesis for small values of λ or small values of the right hand side of equation 46. The difficulty is that we not know the distribution of the right hand side of equation 46. Note that we can write it in terms of estimated residuals as

$$\begin{aligned}
\lambda &= \left(\frac{(y - X\beta^c)' (y - X\beta^c)}{(y - X\hat{\beta})' (y - X\hat{\beta})} \right)^{-\frac{n}{2}} \\
&= \left(\frac{u' u}{e' e} \right)^{-\frac{n}{2}} \\
e &= y - X\hat{\beta} \\
u &= y - X\beta^c
\end{aligned} \tag{47}$$

This can then be written as

$$\begin{aligned}
\lambda &= \left(\frac{u' u}{e' e} \right)^{-\frac{n}{2}} \\
&= \left(\frac{SSE(\beta^c)}{SSE(\hat{\beta})} \right)^{-\frac{n}{2}}
\end{aligned} \tag{48}$$

So we reject the null hypothesis that $R\beta = r$ if

$$\begin{aligned}
\lambda &= \left(\frac{SSE(\beta^c)}{SSE(\hat{\beta})} \right)^{-\frac{n}{2}} \leq c \\
\Rightarrow \lambda^{-\frac{2}{n}} &= \left(\frac{SSE(\beta^c)}{SSE(\hat{\beta})} \right) \geq c^{-\frac{2}{n}}
\end{aligned} \tag{49}$$

Now subtract $\frac{SSE(\hat{\beta})}{SSE(\hat{\beta})} = 1$ from both sides of equation 49 and simplify

$$\begin{aligned}
 \lambda^{-\frac{2}{n}} &= \frac{SSE(\beta^c)}{SSE(\hat{\beta})} \geq c^{-\frac{2}{n}} \\
 \Rightarrow \lambda^{-\frac{2}{n}} - \frac{SSE(\hat{\beta})}{SSE(\hat{\beta})} &= \frac{SSE(\beta^c)}{SSE(\hat{\beta})} - \frac{SSE(\hat{\beta})}{SSE(\hat{\beta})} \geq c^{-\frac{2}{n}} - \frac{SSE(\hat{\beta})}{SSE(\hat{\beta})} \\
 \Rightarrow \lambda^{-\frac{2}{n}} - 1 &= \frac{SSE(\beta^c) - SSE(\hat{\beta})}{SSE(\hat{\beta})} \geq c^{-\frac{2}{n}} - 1
 \end{aligned} \tag{50}$$

Now multiply by $\frac{n-k}{m}$ to obtain

$$\begin{aligned}
 \lambda^{-\frac{2}{n}} - 1 &= \frac{SSE(\beta^c) - SSE(\hat{\beta})}{SSE(\hat{\beta})} \geq c^{-\frac{2}{n}} - 1 \\
 \Rightarrow \frac{n-k}{m} \left(\lambda^{-\frac{2}{n}} - 1 \right) &= \frac{n-k}{m} \left(\frac{SSE(\hat{\beta}) - SSE(\beta^c)}{SSE(\hat{\beta})} \right) \geq \frac{n-k}{m} \left(c^{-\frac{2}{n}} - 1 \right)
 \end{aligned} \tag{51}$$

We reject the null hypothesis if the value in equation 51 is greater than some arbitrary value

$$b = \left(\frac{n-k}{m} \right) \left(c^{-\frac{2}{n}} - 1 \right).$$

The question is then finding the distribution of the test statistic in equation 51. We can also write the test statistic as

$$\frac{n-k}{m} \left(\lambda^{-\frac{2}{n}} - 1 \right) = \left(\frac{\frac{SSE(\beta^c) - SSE(\hat{\beta})}{m}}{\frac{SSE(\hat{\beta})}{n-k}} \right) \tag{52}$$

We have already shown that the numerator in equation 52 is the same as the numerator in equations 37-39. Therefore this statistic is equivalent to the those statistics and is distributed as an F. Specifically,

$$\frac{n-k}{m} \left(\lambda^{-\frac{2}{n}} - 1 \right) = \left(\frac{\frac{SSE(\beta^c) - SSE(\hat{\beta})}{m}}{\frac{SSE(\hat{\beta})}{n-k}} \right) \sim F(m, n-k) \tag{53}$$

Therefore the likelihood ratio test and the F test for a set of linear restrictions $R\beta = r$ in the classical normal linear regression model are equivalent.

c. asymptotic distribution of the likelihood ratio statistic

We show in section on non-linear estimation that

$$-2\ln\lambda(x_1, x_2, \dots, x_n) = -2\ln\left(\frac{\sup_{\theta \in H_0} L(\theta; x_1, x_2, \dots, x_n)}{\sup_{\theta \in H_0 \cup H_a} L(\theta; x_1, x_2, \dots, x_n)}\right) = -2(\ell^c - \ell) \stackrel{d}{\sim} \chi^2(m) \quad (54)$$

where ℓ^c is the value of the log-likelihood function for the model estimated subject to the null hypothesis, ℓ is the value of the log-likelihood function for the unconstrained model, and there are m restrictions on the parameters in the form $R(\theta) = r$.

6. Some examples

a. two linear constraints

Consider the unconstrained model

$$y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + \varepsilon_t \quad (55)$$

with the usual assumptions. Consider the null hypothesis $H_0: \beta_2 = \beta_3 = 0$ which consists of two restrictions on the coefficients. We can test this hypothesis by running two regressions and then forming the test statistic.

1) Estimate

$$y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + \varepsilon_t \quad (56)$$

and obtain $SSE = (n - 4)s^2$ where s^2 is the estimated variance from the unrestricted model.

2) Estimate

$$y_t = \beta_1 + \beta_4 x_{t4} + \varepsilon_t \quad (57)$$

and obtain $SSE^c = (n - 2)s^2$ where s^2 is the estimated variance from the unrestricted model.

3) Construct the test statistic

$$\begin{aligned} F(m, n - k) &= \frac{\frac{SSE(\hat{\beta}^c) - SSE(\hat{\beta})}{(n - k)^c - (n - k)}}{\frac{SSE(\hat{\beta})}{n - k}} = \frac{\frac{SSE(\hat{\beta}^c) - SSE(\hat{\beta})}{2}}{\frac{SSE(\hat{\beta})}{n - 4}} \\ &= \left(\frac{n - 4}{2} \right) \left(\frac{SSE(\hat{\beta}^c) - SSE(\hat{\beta})}{SSE(\hat{\beta})} \right) \end{aligned} \quad (58)$$

- b. equality of coefficients in two separate regressions (possibly different time periods)

We sometimes want to test the equality of the full set of regression coefficients in two regressions. Consider the two models below

$$\begin{aligned} y^1 &= X^1 \beta^1 + \varepsilon^1, \quad n_1 \text{ observations} \\ y^2 &= X^2 \beta^2 + \varepsilon^2, \quad n_2 \text{ observations} \end{aligned} \quad (59)$$

We may want to test the hypothesis $H_0: \beta^1 = \beta^2$ where there are k coefficients and thus k restrictions. Rewrite the model in stacked form as

$$y = \begin{pmatrix} y^1 \\ y^2 \end{pmatrix} = \begin{pmatrix} X^1 & 0 \\ 0 & X^2 \end{pmatrix} \begin{pmatrix} \beta^1 \\ \beta^2 \end{pmatrix} + \begin{pmatrix} \varepsilon^1 \\ \varepsilon^2 \end{pmatrix} \quad (60)$$

and estimate as usual to obtain SSE (no restriction). Note that $(n - k) = n_1 + n_2 - 2k$. Then impose the restriction (hypothesis that $\beta^1 = \beta^2 = \beta$) by writing the model in equation 60 as

$$y = \begin{pmatrix} y^1 \\ y^2 \end{pmatrix} = \begin{pmatrix} X^1 \\ X^2 \end{pmatrix} \beta + \begin{pmatrix} \varepsilon^1 \\ \varepsilon^2 \end{pmatrix} \quad (61)$$

Estimate equation 61 using least squares to obtain the constrained sum of squared errors SSE^c with degrees of freedom $(n - k)^c = n_1 + n_2 - k$. Then construct the test statistics

$$\begin{aligned} F(k, n_1 + n_2 - 2k) &= \frac{\frac{SSE(\hat{\beta}^c) - SSE(\hat{\beta})}{(n - k)^c - (n - k)}}{\frac{SSE(\hat{\beta})}{n - k}} = \frac{\frac{SSE(\hat{\beta}^c) - SSE(\hat{\beta})}{k}}{\frac{SSE(\hat{\beta})}{n_1 + n_2 - 2k}} \\ &= \left(\frac{n_1 + n_2 - 2k}{k} \right) \left(\frac{SSE(\hat{\beta}^c) - SSE(\hat{\beta})}{SSE(\hat{\beta})} \right) \end{aligned} \quad (62)$$

C. Forecasting

1. introduction

Let

$$y_t = F(x_t, \beta) + \varepsilon_t \quad (63)$$

denote the stochastic relationship between the variable y_t and the vector of variables $x_t = (x_{t1}, \dots, x_{tk})$. β represents a vector of parameters. Forecasts are generally made by estimating the vector of parameters $\hat{\beta}$, determining the appropriate vector x_t (sometimes forecasted by \hat{x}_t) and then evaluating

$$\hat{y}_t = F(\hat{x}_t, \hat{\beta}) \quad (64)$$

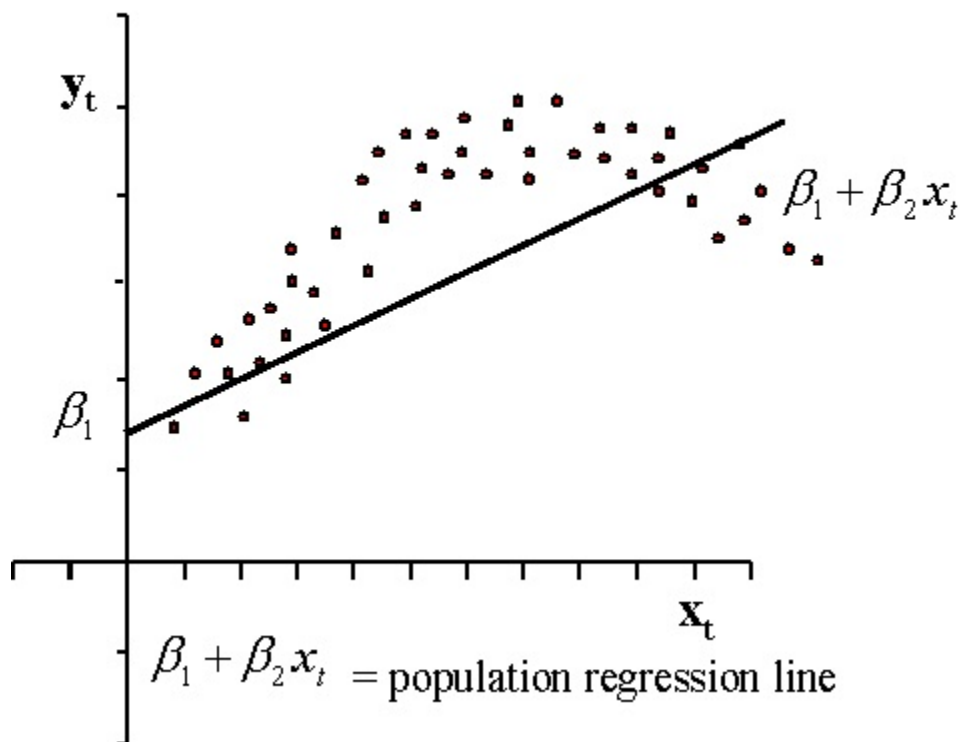
The forecast error is given by

$$fe = y_t - \hat{y}_t \quad (65)$$

There are at least four factors which contribute to forecast error.

a. incorrect functional form

Incorrect Functional Form



- b. the existence of the random disturbance ε_t

Even if the "appropriate" value of x_t were known with certainty and the parameters β were known, the forecast error would not be zero because of the presence of ε_t . The forecast error is given by

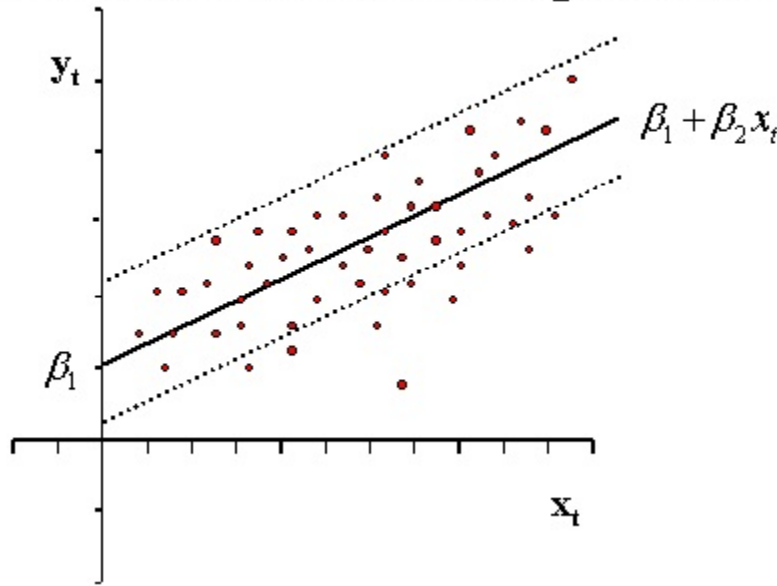
$$\begin{aligned} f_e(\beta \text{ known}) &= y_t - \hat{y}_t(\beta \text{ known}) = y_t - F(x_t, \beta) = \varepsilon_t \\ \sigma_{f_e}^2(\beta \text{ known}) &= \text{Variance}(f_e) \\ &= \text{Var}(\varepsilon_t) = \sigma^2. \end{aligned} \quad (66)$$

Confidence intervals for y_t would be obtained from

$$\Pr[F(x_t, \beta) - t_{\alpha/2}\sigma < y_t < F(x_t, \beta) + t_{\alpha/2}\sigma] = 1 - \alpha \quad (67)$$

We can see this graphically as

Confidence Intervals for Regression Line



- c. uncertainty about β

Consider a set of n_0 observations not included in the original sample of n observations. Let X_0 denote the n_0 observations on the regressors and y_0 the observations on y . The relevant model for this out of sample forecast is

$$y_0 = X_0 \beta + \varepsilon_0 \quad (68)$$

where $E(\varepsilon_0) = 0$ and $E(\varepsilon_0 \varepsilon_0') = \sigma^2 I_{n_0}$ and ε_0 is independent of ε from the sample period.

Now consider a forecast for y_0 given by

$$\hat{y}_0 = X_0 \hat{\beta} \quad (69)$$

where $\hat{\beta}$ is the OLS estimator obtained using the initial n observations. Then let v_0 be the set of forecast errors defined by

$$v_0 = y_0 - X_0 \hat{\beta} = y_0 - X_0 (X'X)^{-1} X'y \quad (70)$$

We can show that v_0 has an expectation of zero as follows

$$E(v_0) = E(y_0 - X_0 \hat{\beta}) = X_0 \beta - X_0 \hat{\beta} = 0 \quad (71)$$

Now consider the variance of the forecast error. We can derive it as

$$\begin{aligned} E(v_0 v_0') &= E(y_0 - X_0 \hat{\beta}) (y_0 - X_0 \hat{\beta})' \\ &= E(X_0 \beta + \varepsilon_0 - X_0 \hat{\beta}) (X_0 \beta + \varepsilon_0 - X_0 \hat{\beta})' \\ &= E(X_0(\beta - \hat{\beta}) + \varepsilon_0) (X_0(\beta - \hat{\beta}) + \varepsilon_0)' \\ &= E(X_0(\hat{\beta} - \beta) - \varepsilon_0) (X_0(\hat{\beta} - \beta) - \varepsilon_0)' \\ &= E(X_0(\hat{\beta} - \beta) - \varepsilon_0) \left((\hat{\beta} - \beta)' X_0' - \varepsilon_0' \right) \\ &= E \left(X_0(\hat{\beta} - \beta)(\hat{\beta} - \beta)' X_0' - \varepsilon_0(\hat{\beta} - \beta)' X_0' - X_0(\hat{\beta} - \beta) \varepsilon_0' + \varepsilon_0 \varepsilon_0' \right) \end{aligned} \quad (72)$$

Now note that ε_0 and $\hat{\beta} - \beta$ are independent by the independence of ε_0 and ε and the independence of ε and $\hat{\beta} - \beta$ which we proved earlier in the section on statistical inference (equations 59-62). As a result the middle two terms in (72) will have an expectation of zero. We then obtain

$$\begin{aligned} E(v_0 v_0') &= E \left(X_0(\hat{\beta} - \beta)(\hat{\beta} - \beta)' X_0' - \varepsilon_0(\hat{\beta} - \beta)' X_0' - X_0(\hat{\beta} - \beta) \varepsilon_0' + \varepsilon_0 \varepsilon_0' \right) \\ &= X_0 E \left((\hat{\beta} - \beta)(\hat{\beta} - \beta)' \right) X_0' + E(\varepsilon_0 \varepsilon_0') \\ &= X_0 \sigma^2 (X'X)^{-1} X_0' + \sigma^2 I \\ &= \sigma^2 \left(X_0 (X'X)^{-1} X_0' + I \right) \end{aligned} \quad (73)$$

This indicates the sampling variability is composed of two parts, that due to the equation error, ε_0 , and that due to the error in estimating the unknown parameters. We should note that

$\hat{y}_0 = X_0 \hat{\beta}$ can be viewed as an estimator of $E(y_0)$ and as a predictor of y_0 . In other words

$\hat{y}_0 = X_0 \hat{\beta}$ is the best predictor we have of the regression line, and of an individual y_0 . The least squares estimator of $E(y_0)$ is $X_0 \hat{\beta}$ which has expected value $X_0 \beta$. Now consider the covariance

matrix for the random vector $(X_0\hat{\beta} - X_0\beta)$. To simplify the derivation write it as $(X_0\hat{\beta} - X_0\beta) = X_0(\hat{\beta} - \beta)$. Then we can compute this covariance as

$$\begin{aligned} E(X_0(\hat{\beta} - \beta)(\hat{\beta} - \beta)'X_0') &= X_0E((\hat{\beta} - \beta)(\hat{\beta} - \beta)')X_0' \\ &= X_0\sigma^2(X'X)^{-1}X_0' \\ &= \sigma^2(X_0(X'X)^{-1}X_0') \end{aligned} \quad (74)$$

This is less than the covariance in equation 73 by the variance of the equation error σ^2I . Now consider the case where $n_0 = 1$ and we are forecasting a given y_0 for a given x_0 , where x_0' is a row vector. Then the predicted value of y_0 for a given value of x_0 is given by

$$\hat{y}_0 = x_0' \hat{\beta} \quad (75)$$

The prediction error is given by

$$\text{prediction error} = v_0 = y_0 - \hat{y}_0 \quad (76)$$

The variance of the prediction error is given by

$$\text{Var}(y_0 - \hat{y}_0) = \sigma^2(x_0'(X'X)^{-1}x_0 + 1) \quad (77)$$

The variance of $E(y_0 | x_0)$ is

$$\text{Var}(\hat{y}_0 - X_0\beta) = \sigma^2(x_0'(X'X)^{-1}x_0) \quad (78)$$

Based on these variances we can consider confidence intervals for y_0 and $E(y_0 | x_0)$ where we estimate σ^2 with s^2 . The confidence interval for $E(y_0 | x_0)$ is

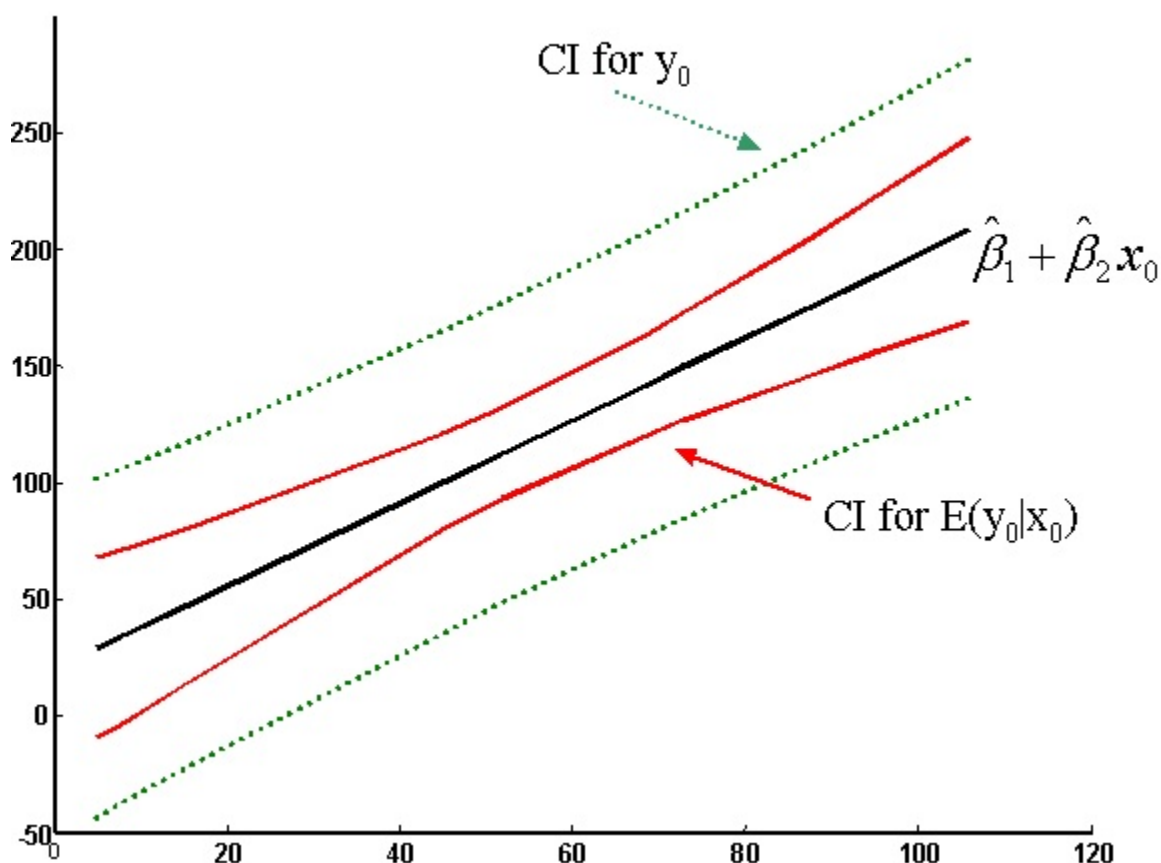
$$\Pr[x_0' \hat{\beta} - t_{\alpha/2} s_{\hat{y}_0} < y_0 < x_0' \hat{\beta} + t_{\alpha/2} s_{\hat{y}_0}] = 1 - \alpha \quad (79)$$

where $s_{\hat{y}_0}$ is the square root of the variance in equation 78. The confidence interval for y_0 is given by

$$\Pr[x_0' \hat{\beta} - t_{\alpha/2} s_{fe} < y_0 < x_0' \hat{\beta} + t_{\alpha/2} s_{fe}] = 1 - \alpha \quad (80)$$

where s_{fe} is the square root of the variance in equation 77. Graphically the confidence bounds in predicting an individual y_0 are wider than in predicting the expected value of y_0 .

Confidence Intervals for Prediction



- d. uncertainty about X . In many situations the value of the independent variable also needs to be predicted along with the value of y . A "poor" estimate of x_t will likely result in a poor forecast for y .
- e. predictive tests

One way to consider the accuracy of a model is to estimate it using only part of the data. Then use the remaining observations as a test of the model.

- 1) Compute $\hat{\beta}$ using observations 1 to n_1 .
- 2) Compute $\hat{y} = X\hat{\beta}$ using the observations on the x 's from n_1 to n .
- 3) Compare the predicted values of y with the actual ones from the sample for observations n_1 to n .

Literature Cited

Chow, G. C., "Tests of Equality Between Subsets of Coefficients in Two Linear Regressions," Econometrica, 28(1960), 591-605.

Fisher, F. M., "Tests of Equality Between Sets of Coefficients in Two Linear Regressions: An Expository Note," Econometrica, 38(1970), 361-66.

That equation 38 is distributed as an F random variable can also be seen by remembering that an F is the ratio of two chi-square random variable, each divided by its degrees of freedom. We have previously shown (equation 177 in the section on Statistical Inference) that $SSE(\hat{\beta})$ is distributed as a $\chi^2(n-k)$ random variable. Note that the vector $\mathbf{r} - \mathbf{R}\hat{\beta}$ is distributed normally with a mean of zero. Its variance is given by

$$\begin{aligned}
 Var(\mathbf{r} - \mathbf{R}\hat{\beta}) &= \mathbf{R}Var(\hat{\beta})\mathbf{R}' \\
 &= \mathbf{R}\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' \\
 &= \sigma^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'
 \end{aligned}
 \tag{81}$$