Computation and Distribution of Constrained Estimator:

Consider the null hypothesis H_0: $R\beta = r$, where R is $q \times k$ and r is $q \times 1$. We suppose there are genuinely q restrictions under H_0, so rank $(R) = q$.

Let $\hat{\beta}$ be the unconstrained estimator,

i.e., $\hat{\beta} = (X'X)^{-1}X'y$.

Let b be the constrained estimator satisfying $Rb = r$. (Typically, $R\hat{\beta} \neq r$.)

Proposition:
$b = \hat{\beta} + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(r - R\hat{\beta})$

Proof:
Let $S(\tilde{b}) = (y - X\tilde{b})'(y - X\tilde{b}) - 2\lambda(R\tilde{b} - r)$.
The constrained estimator b satisfies the first order conditions (2's cancel):
(1) $\quad -X'y + X'Xb - R'\lambda = 0$
(2) $\quad Rb - r = 0$
Thus $b = \hat{\beta} + (X'X)^{-1}R'\lambda$

Let's eliminate λ:
$Rb = R\hat{\beta} + R(X'X)^{-1}R'\lambda$
Since $Rb = r$,
$[R((X'X)^{-1}R']^{-1}r = [R(X'X)^{-1}R']^{-1}R\hat{\beta} + \lambda$.

Thus, $\lambda = [R(X'X)^{-1}R']^{-1}(r - R\hat{\beta})$.
Substitute out λ in the definition of b:
$b = \hat{\beta} + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(r - R\hat{\beta})$ ■

Sampling distribution of b

First step is to find the mean and variance of b:

Proposition: $Eb = \beta$. (Under H_0)

Proof :Substitute $\hat{\beta}$ in the definition of b:

$$
\begin{aligned}
b = \beta &+ (X'X)^{-1}X'\varepsilon \\
&+(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}[r - R\beta - R(X'X)^{-1}X'\varepsilon] \\
&= \beta + [I - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R](X'X)^{-1}X'\varepsilon,
\end{aligned}
$$

using $r = R\beta$.

From this we see that $Eb = \beta$. ■

Proposition: $V(b) \leq V(\hat{\beta})$.

Proof: Let $A = R(X'X)^{-1}R'$.

Note that:

$$b - \beta = [I - (X'X)^{-1}R'A^{-1}R](X'X)^{-1}X'\varepsilon.$$

$$\begin{aligned} V(b) &= E(b-\beta)(b-\beta)' \\ &= \sigma^2[I - (X'X)^{-1}R'A^{-1}R](X'X)^{-1} \\ &\qquad [I - (X'X)^{-1}R'A^{-1}R]', \\ &\qquad\qquad \text{since } E\varepsilon\varepsilon' = \sigma^2 I \\ &= \sigma^2[(X'X)^{-1} - 2(X'X)^{-1}R'A^{-1}R(X'X)^{-1} \\ &\quad +(X'X)^{-1}R'A^{-1}R(X'X)^{-1}R'A^{-1}R(X'X)^{-1}] \end{aligned}$$

Using the definition of A, this becomes

$$\begin{aligned} V(b) &= \sigma^2[(X'X)^{-1} - (X'X)^{-1}R'A^{-1}R(X'X)^{-1}] \\ &\leq V(\hat{\beta}) = \sigma^2(X'X)^{-1} \qquad \text{(why?)} \ \blacksquare \end{aligned}$$

- What is the relation to the Gauss-Markov theorem?
- Why doesn't this expression depend on r?

Proposition: Under normality, we have the complete sampling distribution of b with the mean and the variance calculated above.

Estimation of σ^2:

- What is the unbiased estimator under restriction?
- What is the ML estimator?

F Tests

Let e and e^* be the vector of restricted and unrestricted residuals respectively.

Proposition:
$e'e - e^{*\prime}e^* = (r - R\hat{\beta})'[R(X'X)^{-1}R']^{-1}(r - R\hat{\beta})$

Proof: $e = y - Xb = y - X\hat{\beta} - X(b - \hat{\beta})$
$= e^* - X(b - \hat{\beta})$
$\Rightarrow e'e = e^{*\prime}e^* + (b - \hat{\beta})'X'X(b - \hat{\beta})$
$\Rightarrow e'e - e^{*\prime}e^* = (r - R\hat{\beta})'[R(X'X)^{-1}R']^{-1}(r - R\hat{\beta})$ ■

Example: Consider $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$ with the restriction $\beta_1 + \beta_2 = 2$. If we substitute for β_1, we get

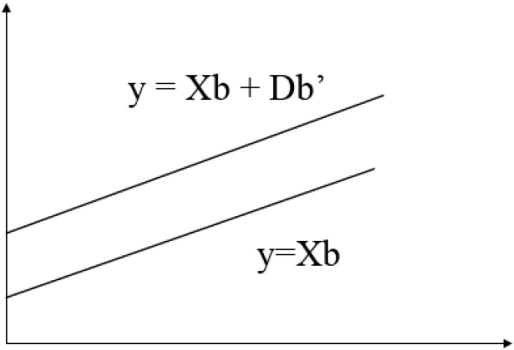
$$y = \beta_0 + (2 - \beta_2)x_1 + \beta_2 x_2 + \varepsilon$$
$$y = \beta_0 + 2x_1 - \beta_2 x_1 + \beta_2 x_2 + \varepsilon$$
$$\Rightarrow y - 2x_1 = \beta_0 + \beta_2(x_2 - x_1) + \varepsilon$$

- Regress $(y - 2x_1)$ on a constant term and $(x_2 - x_1)$, and get the sum squared residuals from this restricted regression $(e'e)$.
- Regress y on a constant term, x_1 and x_2, and get the sum squared residuals from this unrestricted regression $(e^{*\prime}e^*)$.
- Compare the sums of squared residuals from these regressions.

Dummy Variables

Here we define a new variable D equal to 0 or 1 indicating absence or presence of a characteristic.

This allows the intercept to differ.



Example: homeowners/renters, male/female, regulation applies/regulation doesn't apply, etc.

Dummy variable trap

Suppose:

$$D_1 = \begin{cases} 1 & \text{if a characteristic is present} \\ 0 & \text{if a characteristic is absent} \end{cases}$$

$$D_0 = \begin{cases} 1 & \text{if a characteristic is absent} \\ 0 & \text{if a characteristic is present} \end{cases}$$

then $D_1 + D_0 = 1$ and there is a problem of perfect collinearity if the regressors include $X_1 = 1 \in R^n \Longrightarrow X'X$ is be not invertible.
Solutions: drop either $1 \in R^n$ or one between D_1 and D_0.

Two alternative reparametrizations

Consider the model

$$Y = D_1\beta_1 + D_0\beta_0 + X^*\beta^* + \varepsilon$$

Given that

$$D_1\beta_1 + D_0\beta_0 = D_1\beta_1 + (1 - D_1)\beta_0 = \beta_0 + D_1\underbrace{(\beta_1 - \beta_0)}_{\beta_1^*}$$

the model could be reparametrized as

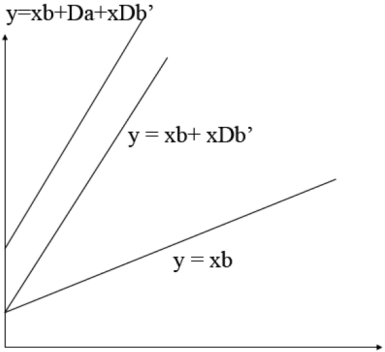
$$Y = \beta_0 + D_1\beta_1^* + X^*\beta^* + \varepsilon$$

Interpretation of coefficients:

- β_1 is the conditional expectation of variable y when $X^* = 0$ and the characteristic is present;
- β_0 is the conditional expectation of variable y when $X^* = 0$ and the characteristic is absent;
- $\beta_1^* = (\beta_1 - \beta_0)$ is the average difference of y between the units for which the characteristic is present and those for which the characteristic is absent.

A test for $H_0 : \beta_1^* = 0$ is then a test for a null average effect of the characteristic on variable y.

Interactions with continuous regressors:



Example

Suppose education is reported in grouped form:
0-8 years; 9-12years; 12+ years

How should we set up the dummy variables?

One temptation is to code

$d = 0$ if 0-8 years of educaction
$= 1$ if 9-12 years of eduction
$= 2$ if 12+ years of education

This is very restrictive and probably unsound.

A better set up would be to use 2 dummies:

$$\begin{aligned} d_1 &= 1 \text{ if 0-8 years of education} \\ &= 0 \text{ else} \\ d_2 &= 1 \text{ if 9-12 years of education} \\ &= 0 \text{ else} \end{aligned}$$

The first set up imposes that the effect of having 12+ years of education is twice the effect of having 9-12 years of education. In general, class variables with several classes require many dummies.

Practical matters:

Often you will run across categorical variables - with no natural ordering. It is usually appropriate to do a fequency distribution and form dummy variables on that basis.

For example, suppose the variable is color, and you have out of a sample of 100; 25 red, 5 yellow, 40 blue, 1 green, 4 purple, etc. (small numbers for the remaining colors).

It is probably appropriate to make a dummy for red, one for blue, and use "other" as the base.

Plotting residuals, especially for the "base" observations, will tell you if this fails.

Multicollinearity

The problem is lack of data information when $X'X$ is singular (recall picture) or "nearly" singular.

If some X's move together, it is difficult to sort their separate effects on y. More data *does* help.

Other sources of information are useful. Purely "technical" remedies for collinearity work by imposing arbitrary and sometimes hidden "information". Never use ridge regression in an economic application.

The problem of multicollinearity in K-variable regression is equivalent to the problem of small sample size in estimating a mean.