Economics 620, Lecture 6:

More on the K-Variable Linear Model

Computation and Distribution of Constrained Estimators:

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:

$$b = \beta + (X'X)^{-1}X'\varepsilon$$

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

$$= \beta + [I - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R](X'X)^{-1}X'\varepsilon,$$

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:

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

$$V(b) = E(b-\beta)(b-\beta)'$$

$$= \sigma^2[I - (X'X)^{-1}R'A^{-1}R](X'X)^{-1}$$

$$[I - (X'X)^{-1}R'A^{-1}R]',$$

since $E\varepsilon\varepsilon' = \sigma^2 I$

$$= \sigma^2[(X'X)^{-1} - 2(X'X)^{-1}R'A^{-1}R(X'X)^{-1}$$

$$+(X'X)^{-1}R'A^{-1}R(X'X)^{-1}R'A^{-1}R(X'X)^{-1}]$$

Using the definition of A, this becomes

$$V(b) = \sigma^2[(X'X)^{-1} - (X'X)^{-1}R'A^{-1}R(X'X)^{-1}]$$

$$\leq V(\hat{\beta}) = \sigma^2(X'X)^{-1}$$ (why?) ■

- 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 $(e'e)$.

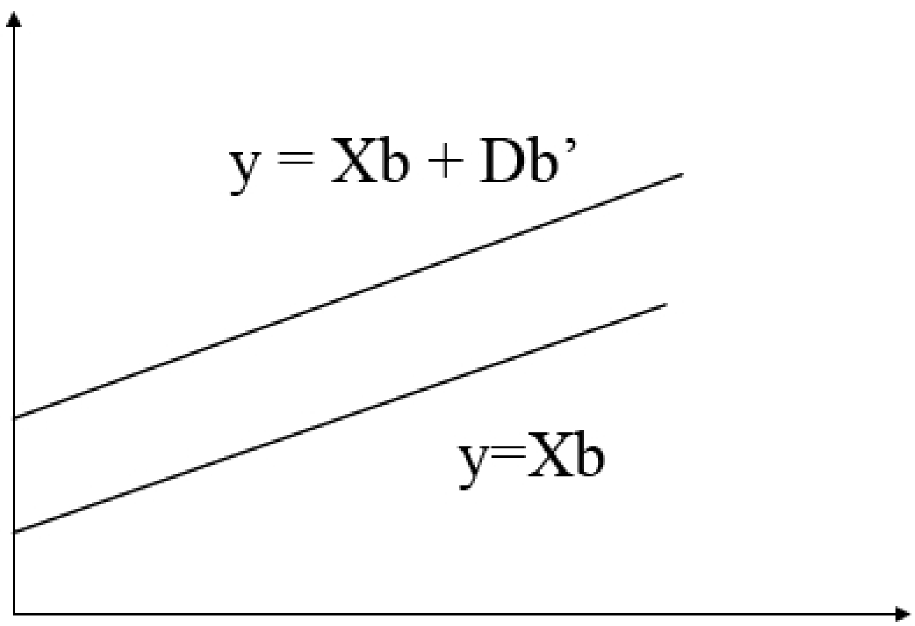
- Regress y on a constant term, x_1 and x_2, and get the sum squared residuals from this unrestricted regression $(e^{*\prime}e^*)$.

- 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 $X_2 = 1$ if the characteristic is present

$= 0$ if the characteristic is not present

and $X_3 = 1$ if the characteristic is not present

$= 0$ if the characteristic is present.

Then $X_2 + X_3 = 1 = X_1$ if the regression contains the constant term $X_1 = 1 \in R^N$. .... *And*?

Interactions between dummies for different characteristics:

Suppose X_2 is the dummy variable for characteristic 1 and X_3 is the dummy variable for characteristic 2. Let $X_4 = X_2 * X_3$ (elementwise).

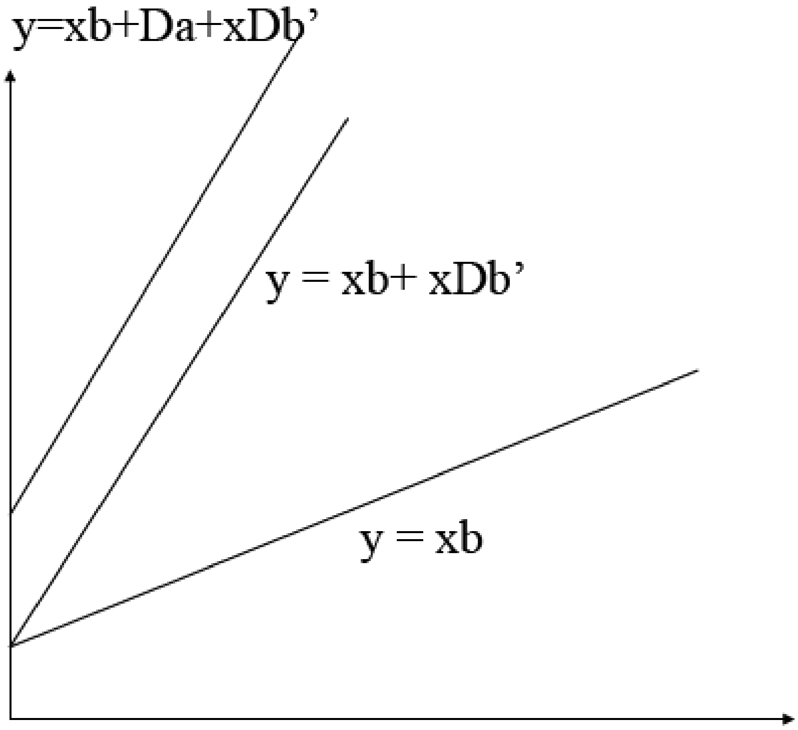
That is, $X_4 = 1$ if both characteristics are present.

$= 0$ if only one or none of the characteristics is present.

Then the (marginal) effect of characteristic 1 is β_2; effect of charactersitic 2 is β_3; effect of *both* is $\beta_2 + \beta_3 + \beta_4$.

This could be set up differently. Although different set ups will give different coefficients, correct interpretation of these coefficients will give the same estimated effects.

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:

$d_1 = 1$ if 0-8 years of education

$\quad = 0$ else

$d_2 = 1$ if 9-12 years of education

$\quad = 0$ else

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.

Micronumerosity

Goldberger gives an example that puts the problem in perspective.

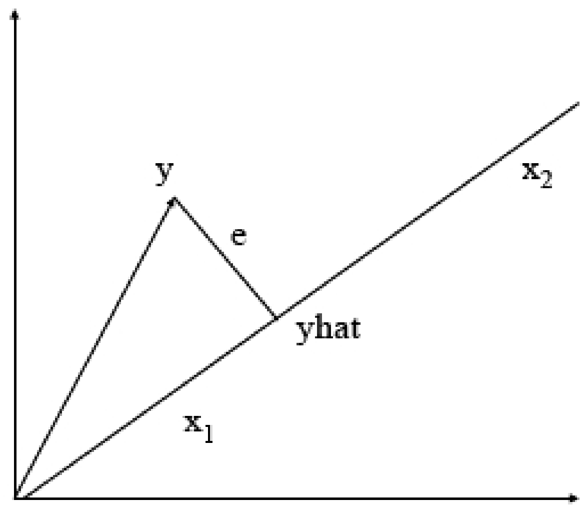
Consider estimating a normal mean μ. The usual estimator is the sample mean with variance σ^2/N. This is a regression model, $Ey = \mu \mathbf{1}_N$, $V(y) = \sigma^2 I_N$.

When N is small, the sampling variance is large - "micronumerosity".

Extreme micronumerosity occurs when $N = 0$. Of course, this is just multicolinearity, since $X'X = N$ in the regression interpretation, and $X'X$ singular is $N = 0$.

Near multicolinearity corresponds to small N.

Multicollinearity: effect on $\hat{y}$??



Example with $n = 2$ and x_1 and x_2 collinear. What happens in the full rank case?

