0.1 Partialling Out Interpretation of OLS regression

Recall:

$$Y = X\beta + \varepsilon$$

Now lets call a subset of variables X_{s1} and another (non intersecting) subset of variables as X_{s2}. In other words, X is partitioned into X_{s1} and X_{s2}. We can therefore write:

$$Y = X_{s1}\beta_{s1} + X_{s2}\beta_{s2} + \varepsilon$$

where β is appropriately partitioned into β_{s1} and β_{s2}.

Recall the normal equations:

$$X'X\overset{\wedge}{\beta} = X'Y \tag{1}$$

(for the rest of the course, unless specified, all estimators are OLS).

(1) can be rewritten as

$$\begin{pmatrix} X'_{s1}X_{s1} & X'_{s1}X_{s2} \\ X'_{s2}X_{s1} & X'_{s2}X_{s2} \end{pmatrix} \begin{pmatrix} \overset{\wedge}{\beta}_{s1} \\ \overset{\wedge}{\beta}_{s2} \end{pmatrix} = \begin{pmatrix} X'_{s1}Y \\ X'_{s2}Y \end{pmatrix}$$

The First Line implied by this condition:

$$X'_{s1}X_{s1}\overset{\wedge}{\beta}_{s1} + X'_{s1}X_{s2}\overset{\wedge}{\beta}_{s2} = X'_{s1}Y$$

or

$$\overset{\wedge}{\beta}_{s1} = \left(X'_{s1}X_{s1}\right)^{-1} X'_{s1}Y - \left(X'_{s1}X_{s1}\right)^{-1} X'_{s1}X_{s2}\overset{\wedge}{\beta}_{s2} \tag{2}$$

1. If two sets of variables are orthogonal, that is, $X'_{s1}X_{s2} = 0$, then running the regression with all variables or seperately with each set of results yields the same $\overset{\wedge}{\beta}$ s.

This has important implications on the impact of variables that have been left out of the regresion.

The Second Line implied by this condition:

$$X'_{s2}X_{s1}\overset{\wedge}{\beta}_{s1} + X'_{s2}X_{s2}\overset{\wedge}{\beta}_{s2} = X'_{s2}Y$$

Substitute the value of $\overset{\wedge}{\beta}_{s1}$ from (2):

$$X'_{s2}X_{s1}\left[\left(X'_{s1}X_{s1}\right)^{-1} X'_{s1}Y - \left(X'_{s1}X_{s1}\right)^{-1} X'_{s1}X_{s2}\overset{\wedge}{\beta}_{s2}\right]$$

$$+X'_{s2}X_{s2}\overset{\wedge}{\beta}_{s2} = X'_{s2}Y$$

This implies

$$X'_{s2}X_{s2}\overset{\wedge}{\beta}_{s2} - X'_{s2}X_{s1}\left(X'_{s1}X_{s1}\right)^{-1} X'_{s1}X_{s2}\overset{\wedge}{\beta}_{s2}$$

$$= X'_{s2}Y - X'_{s2}X'_{s1}\left(X'_{s1}X'_{s1}\right)^{-1}X'_{s1}Y$$

or,

$$X'_{s2}\left[I - X_{s1}\left(X'_{s1}X_{s1}\right)^{-1}X'_{s1}\right]X_{s2}\overset{\wedge}{\beta}_{s2}$$

$$= X'_{s2}\left[I - X_{s1}\left(X'_{s1}X_{s1}\right)^{-1}X'_{s1}\right]Y$$

Recall the definition of the residual matrix: Therefore, we can re write this expression in terms of the residual matrix:

$$X'_{s2}M_{X_{s1}}X_{s2}\overset{\wedge}{\beta}_{s2} = X'_{s2}M_{X_{s1}}Y$$

Hence:

$$\overset{\wedge}{\beta}_{s2} = \left(X'_{s2}M_{X_{s1}}X_{s2}\right)^{-1}X'_{s2}M_{X_{s1}}Y$$

Recall again, that $M_{X_{s1}}$ is idempotent: So, $M_{X_{s1}}.M_{X_{s1}} = M_{X_{s1}}$. Also $M_{X_{s1}} = M'_{X_{s1}}$Therefore,

$$\overset{\wedge}{\beta}_{s2} = \left(X'_{s2}M'_{X_{s1}}.M_{X_{s1}}X_{s2}\right)^{-1}X'_{s2}M_{X_{s1}}.M_{X_{s1}}Y$$

Define

$$X^*_{s2} \equiv M_{X_{s1}}X_{s2}$$

and

$$Y^* \equiv M_{X_{s1}}Y$$

Then

$$\overset{\wedge}{\beta}_{s2} = (X^{*\prime}_{s2}X^*_{s2})^{-1}X^*_{s2}Y^*$$

What is X^*_{s2} ? Its the residual when you run a regression of X_{s2} on X_{s1}. What is Y^*? Its the residual when you run a regression of Y on X_{s1}. Therefore is $\overset{\wedge}{\beta}_{s2}$ is what you would get as the coefficient of X_{s2} if

STEP 1: Run a regression of Y on X_{s1}. calculate and store residuals ε_Y

STEP 2: Run a regression of each x in X_{s2} on X_{s1}. Calculate and store residuals ε_x.

STEP 3: Run a regression of ε_Y with all ε_x s as regressors. The coefficient of each x in X_{s2} will be what you would obtain if you ran a regression of Y on X_{s1} and X_{s2}.

0.2 Goodness of Fit

To evaluate the goodness of fit we look at how much of the variation in Y has been explained. We will look at models with a constant (intercept term). So the algebric properties when we run a model with an interecept term are relevant here.

The varition in Y can be summarized by the Total Sum of Squares (TSS): $\sum_{i=1}^{n}(y_i - \overline{y})^2$.

To represent this in matrix notation, we can use M_i where $i = \begin{pmatrix} 1 \\ 1 \\ : \\ 1 \end{pmatrix}$.

what is $M_i Y$?

$$M_i Y = [I - i(i'i)^{-1}i']Y = [I - i(n)^{-1}i']Y$$

Now,

$$\frac{1}{n}ii'Y = \frac{1}{n}i\sum_{i=1}^{n} y_i = i\overline{y}$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} Y = \begin{pmatrix} y_1 \\ : \\ y_n \end{pmatrix} = Y$$

So,

$$M_i Y = \begin{pmatrix} y_1 \\ : \\ y_n \end{pmatrix} - \begin{pmatrix} \overline{y} \\ : \\ \overline{y} \end{pmatrix} = \begin{pmatrix} y_1 - \overline{y} \\ : \\ y_n - \overline{y} \end{pmatrix}$$

Therefore

$$\sum_{i=1}^{n}(y_i - \overline{y})^2 = (M_i Y)^{'} M_i Y = Y'M_i'M_i Y = Y'M_i Y$$

$$= Y'\left[M_i X\hat{\beta} + M_i\hat{\varepsilon}\right]$$

Since in a model with an intercept term the sum of residuals is zero, therefore,the mean residuals, $\overline{\hat{\varepsilon}} = 0$. Thus $M_i\hat{\varepsilon} = \hat{\varepsilon}$. This implies

$$\sum_{i=1}^{n}(y_i - \overline{y})^2 = Y'\left[M_i X\hat{\beta} + \hat{\varepsilon}\right]$$

Now substituting Y gives us

$$= \left(\hat{\beta}'X' + \hat{\varepsilon}'\right)\left(M_i X\hat{\beta} + \hat{\varepsilon}\right)$$

$$= \hat{\beta}' X' M_i X \hat{\beta} + \hat{\beta}' X' \hat{\varepsilon} + \hat{\varepsilon}' M_i X \hat{\beta} + \hat{\varepsilon}' \hat{\varepsilon}$$

$$= \hat{\beta}' X' M_i X \hat{\beta} + \hat{\varepsilon}' \hat{\varepsilon}$$

since $X'\hat{\varepsilon} = 0$

$\hat{\varepsilon}'\hat{\varepsilon}$ is called the residual sum of squares ($\operatorname{Re} sidSS$: also called square of residuals). $\hat{\beta}' X' M_i X \hat{\beta}$ is the explained sum of squares ($ExplSS$; also called regression sum of squares).

Therefore

$$TSS = ExplSS + \operatorname{Re} sidSS$$

To judge the explanatory power of a model, we define R^2 (coeffecient of determinination).

$$R^2 = \frac{ExplSS}{TSS} = \frac{\hat{\beta}' X' M_i X \hat{\beta}}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = 1 - \frac{\hat{\varepsilon}'\hat{\varepsilon}}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

By its construction, R^2 lies between 0 and 1. But notice this is only true for the model with the intercept term!

What is this expression for a regression with an intercept and an explanatory variable? (HW)

Notice:

$$\hat{\beta}' X' M_i X \hat{\beta} = \hat{Y}' M_i \hat{Y} = \left(M_i \hat{Y} \right)' M_i \hat{Y}$$

$$= \sum_{i=1}^{n} \left(\hat{y}_i - \bar{\hat{y}}_i \right) = \sum_{i=1}^{n} \left(\hat{y}_i - \bar{y}_i \right)$$

(because in a regression with intercept the mean residual is zero)