Economics 671
First Examination

Good Luck!

(**1: 40 Points**)

Consider the regression model (here I am using boldface to denote vectors/matrices for clarity. Capital letters denote matrices as opposed to vectors):

$$y_i = \boldsymbol{x}_{1i}\boldsymbol{\beta}_1 + x_{2i}\beta_2 + u_i. \tag{1}$$

In the above, x_{2i} is just a scalar, while $\boldsymbol{x}_{1i}$ is a vector, with a one in the first position to capture the intercept parameter.

Critically evaluate the following claim and associated three-step procedure:

1. Run a regression of $\boldsymbol{y}$ on $\boldsymbol{X}_1$ and obtain the residuals, say $\hat{\boldsymbol{u}}_1$.
2. Run a regression of $\boldsymbol{x}_2$ on $\boldsymbol{X}_1$ and obtain the residuals, say $\hat{\boldsymbol{u}}_2$
3. Run a regression of $\hat{\boldsymbol{u}}_1$ on $\hat{\boldsymbol{u}}_2$.

CLAIM: The coefficient obtained from the regression in step 3 is the same as $\hat{\beta}_2$, the OLS estimate of β_2 in equation (1).

Do you agree or disagree with the claim that $\hat{\beta}_2$ is reproduced when following this three-step procedure?

SOLUTION

Regressing y on X_1 gives the residuals:

$$\hat{u}_1 = y - X_1\hat{\theta} = y - X_1(X_1'X_1)^{-1}X_1'y = [I - X_1(X_1'X_1)^{-1}X_1]y = M_1 y.$$

Similarly, when regressing x_2 on X_1:

$$\hat{u}_2 = M_1 x_2.$$

So, finally, in step 3 when regressing $\hat{u}_1$ on $\hat{u}_2$, you obtain the coefficient estimate:

$$\begin{aligned} \hat{\gamma} &= (\hat{u}_2'\hat{u}_2)^{-1}\hat{u}_2{}'\hat{u}_1 \\ &= (x_2'M_1'M_1x_2)^{-1}x_2'M_1'M_1y \\ &= (x_2'M_1x_2)^{-1}x_2'M_1y \end{aligned}$$

where the last line follows using properties of M_1. This is the same as $\hat{\beta}_2$ as discussed in the notes, and so the claim is true.

(**2: 40 Points**)

Suppose you ran the following regression:

$$GPA_i = \beta_0 + \beta_1 Male_i + \beta_2 HrsStudy_i + \epsilon_i,$$

where $Male$ represents a Male indicator variable (it equals one if the respondent is male and zero if the respondent is female) and $HrsStudy$ is the average number of hours studied by the student per week.

Your grumpy old supervisor, never pleased with your results (or anything for that matter), suggests that you should estimate the following regression instead:

$$GPA_i = \gamma_0 + \gamma_1 Female_i + \gamma_2 MinStudy_i + u_i,$$

where $Female$ represents a Female indicator variable, and $MinStudy$ is the average number of Minutes studied per week.

(**2a: 30 Points**) Show your grumpy old supervisor how $\hat{\gamma}$ will relate to $\hat{\beta}$. *Note that he will only be satisfied if your answer makes use of linear algebra.* You have come to learn that is all he understands (and the only way you will get full credit on this question)!

SOLUTION

First note that

$$Female = 1 - Male$$

and define

$$X = \begin{bmatrix} 1 & Male_1 & HrsStudy_1 \\ 1 & Male_2 & HrsStudy_2 \\ \vdots & \ddots & \vdots \\ 1 & Male_n & HrsStudy_n \end{bmatrix}, \quad \text{and} \quad Z = \begin{bmatrix} 1 & Female_1 & MinStudy_1 \\ 1 & Female_2 & MinStudy_2 \\ \vdots & \ddots & \vdots \\ 1 & Female_n & MinStudy_n \end{bmatrix}$$

so that the first regression is a model of GPA on X and the second is a model of GPA on Z. We can relate the covariate matrices as follows:

$$Z = XA, \quad \text{where} \quad A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 60 \end{bmatrix}. \Rightarrow X = ZA^{-1}.$$

Noting

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'y \\ &= \left((A^{-1})'Z'ZA^{-1}\right)^{-1}(A^{-1})'Z'y \\ &= A(Z'Z)^{-1}A'(A')^{-1}Z'y \\ &= A(Z'Z)^{-1}Z'y \\ &= A\hat{\gamma} \end{aligned}$$

This implies

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 60 \end{bmatrix} \begin{bmatrix} \hat{\gamma}_0 \\ \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{bmatrix} = \begin{bmatrix} \hat{\gamma}_0 + \hat{\gamma}_1 \\ -\hat{\gamma}_1 \\ 60\hat{\gamma}_2 \end{bmatrix}.$$

(**2b: 10 Points**) Can you also comment on how the variance-covariance matrices of $\hat{\gamma}$ and $\hat{\beta}$ will be related?

SOULTUON As for the variance-covariance matrix:

$$\text{Var}(\hat{\beta}|X) = \text{Var}(A\hat{\gamma}|X) = A\text{Var}(\hat{\gamma}|X)A'$$

which we can write as

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 60 \end{bmatrix} \begin{bmatrix} V_1 & V_{12} & V_{13} \\ V_{12} & V_2 & V_{23} \\ V_{12} & V_{23} & V_3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 60 \end{bmatrix}.$$

Multipling all of this out, you get:

$$\begin{bmatrix} V_1 + V_2 + 2V_{12} & -(V_2 + V_{12}) & 60(V_{13} + V_{23}) \\ -(V_2 + V_{12}) & V_2 & -60V_{23} \\ 60(V_{13} + V_{23}) & -60V_{23} & 3600V_3 \end{bmatrix}.$$

Note that these expresessions make sense, given the relationship between $\hat{\gamma}$ and $\hat{\beta}$ in the previous section.

(**3. 20 Points**)

Suppose you ran the following regression:

$$y = X\beta + \epsilon$$

and obtained the fitted values $\hat{y}$ and residuals $\hat{\epsilon}$. You then decided to take those and estimate the regression:

$$\hat{\epsilon} = \theta\hat{y} + u.$$

Comment on the value of $\hat{\theta}$ that is produced, as well as the R^2 value that is obtained. Justify your answer mathematically.

SOLUTION

Note

$$\hat{\theta} = (\hat{y}'\hat{y})^{-1}\hat{y}'\hat{\epsilon} = (\hat{y}'\hat{y})^{-1}\hat{\beta}'X'\hat{\epsilon} = 0$$

since $X'\hat{\epsilon} = 0$.

In terms of R^2, consider the numerator of its definition (explained sum of squares or ESS), which gives the sum of squared deviations of the fitted falues from the average of the dependent variable ($\bar{\hat{\epsilon}}$). We know

$$\bar{\hat{\epsilon}} = 0$$

provided the initial regression contains an intercept parameter. In addition, the fitted values are all zero, since $\hat{\theta} = 0$. Thus, $R^2 = 0$.