Econometrics Practice

Date: 24/01/2007

OUTLINE

1.Exercises on Strict Exogeneity
2.OLS Normal Equations
3. Sample Mean of Fitted Values
4. Decomposition of Sample Deviation
5. Some Sample Identities
6. Effects of Change in Unit of Measurement

1.**Exercises on Strict Exogeneity**

i.) *Suppose we have the classical linear regression model with the usual assumptions (1-4). Show that*

$$E(\varepsilon_i \varepsilon_j \mid X) = E(\varepsilon_i \vert\, X) \cdot E(\varepsilon_j \mid X)$$

Proof. We have the following model in mind

$$y_i = x_i^T \beta + \varepsilon_i$$

recall that the strict exogeneity assumption implies that $E(\varepsilon_i \vert\, X) = 0\ \forall i$.In fact, both RHS and LHS (4.assumption) are equal to 0. We will exploit the Law of Iterated expectations (Theorem on Page 31, SFN), namely

$$\begin{aligned} E(E(Y \mid \mathcal{F}_{X,Z}) \mid \mathcal{F}_X)) &= E(Y \mid \mathcal{F}_X) \\ \mathcal{F}_X &\subset \mathcal{F}_{X,Z} \end{aligned}$$

Also notice that

$$(\varepsilon_i, x_i) \perp (\varepsilon_j, x_1, ..x_{i-1}, x_{i+1}...x_n)$$

Then

$$\begin{aligned} E(\varepsilon_i \varepsilon_j \mid X) &= {}^{L.i.E} E[E(\varepsilon_i \varepsilon_j \mid \varepsilon_j X) \mid X] = E[\varepsilon_j \ E(\varepsilon_i \vert\, \varepsilon_j X) \mid X] \\ &= {}^{\perp} E[\varepsilon_j \ E(\varepsilon_i \vert\, X_i) \mid X_j] = E(\varepsilon_i \vert\, X_i) \cdot E(\varepsilon_j \mid X_j) \end{aligned}$$

ii.) *Show that in the model* $y_i = x_i^T\beta + \varepsilon_i$,

$$E(\varepsilon_i|X) = 0 \forall i \Leftrightarrow E(y_i|X) = x_i^T\beta.$$

Proof. "$\Rightarrow$" $E(y_i \mid X) = E(x_i^T\beta + \varepsilon_i \mid X) = E(x_i^T\beta|X) + E(\varepsilon_i \mid X) =^{s.e} x_i^T\beta.$

"$\Leftarrow$" $E(y_i \mid X) = x_i^T\beta \Rightarrow y_i = x_i^T\beta + \varepsilon_i$ and $E(\varepsilon_i| \ X) = 0$. Since $\varepsilon_i = y_i - x_i^T\beta$, given hypothesis $\varepsilon_i = y_i - E(y_i \mid X)$, we apply conditional expectation

$$E(\varepsilon_i| \ X) = E(y_i| \ X) - E(E(y_i| \ X) \mid X) = x_i^T\beta - E(x_i^T\beta \mid X) = 0.$$

iii.) *Given the spherical error variance assumption* $(E(\varepsilon_i^2| \ X) = \sigma^2, E(\varepsilon_i\varepsilon_j \mid X) = 0, i \neq j)$ *and strict exogeneity show that*

$$var(\varepsilon_i) = \sigma^2, cov(\varepsilon_i\varepsilon_j \ ) = 0, i \neq j$$

Proof. We will use the variance and covariance formulas

$$\begin{aligned} var(\varepsilon_i) &= E(\varepsilon_i^2) - E(\varepsilon_i)^2 =^{L.t.E} E[\ E(\varepsilon_i^2| \ X)] - E[\ E(\varepsilon_i| \ X)]^2 = \sigma^2 \\ cov(\varepsilon_i, \varepsilon_j) &= E(\varepsilon_i\varepsilon_j) - E(\varepsilon_i)E(\varepsilon_j) =^{L.t.E} E(E(\varepsilon_i\varepsilon_j \mid X)) - E(E(\varepsilon_i| \ X)) \cdot E(E(\varepsilon_j| \ X)) = 0 \end{aligned}$$

2.**OLS Normal Equations**

We have the following model

$$\begin{aligned} y_i &= x_i^T\beta + \varepsilon_i \ \ , \ i = 1...n \\ x_{i_{(K\times1)}} &= \begin{pmatrix} x_{1i} \\ x_{2i} \\ . \\ . \\ x_{Ki} \end{pmatrix}, \ \beta_{(K\times1)} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ \beta_K \end{pmatrix} \end{aligned}$$

in the lecture notes we showed that

$$\begin{aligned} \tilde{\varepsilon}_i &= y_i - x_i^T\tilde{\beta} \\ \mathbf{b}_{OLS} &= \arg\min_{\tilde{\beta}} \ SSR(\tilde{\beta}) \\ SSR(\tilde{\beta}) &: \ \mathbb{R}^K \to \mathbb{R} \\ \hat{\beta}_{OLS} &= \mathbf{b}_{OLS} = (X^TX)^{-1}X^Ty \end{aligned}$$

In order to avoid confusion with notation, it's worth mentioning that the values with tildas are the hypothetical values, the OLS estimate is conventionally denoted by $\hat{\beta}_{OLS}$ (with a hat), but we used the book notation in lecture notes, i.e $\mathbf{b}_{OLS}$.

It is a common practice to regress the dependent variable y on independent variables (sample data) and on a **constant**. (*the intercept*). The constant variable is seen in the regressor matrix X as the first column of ones, so it might be useful to seperate the matrix into the constant variable and the rest of the independent variables as follows

$$\begin{aligned} y &= X\beta + \varepsilon \\ X &= [x_1 \; X_2] \\ x_{1_{(n\times 1)}} &= \mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ . \\ 1 \end{bmatrix} \\ X_{2_{(n\times(k-1))}} &= \begin{bmatrix} x_2 & x_3 & . & . & x_n \end{bmatrix} \end{aligned}$$

from now on we will denote the sample estimate values with hat, i.e. $\hat{\beta}, \hat{\varepsilon}, \hat{y}$.

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_{1_{(1\times 1)}} \\ \hat{\beta}_{2_{(K-1)\times 1}} \end{bmatrix}_{(k\times 1)}$$

Recall that we obtained the following from the normal equations

$$\begin{aligned} X^T y &= X^T X \; \hat{\beta}_{OLS} \\ \begin{bmatrix} \mathbf{1}^T \\ X_2^T \end{bmatrix} \cdot y &= \begin{bmatrix} \mathbf{1}^T \\ X_2^T \end{bmatrix} \cdot \begin{bmatrix} \mathbf{1} & X_2 \end{bmatrix} \cdot \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} \\ \begin{bmatrix} \mathbf{1}^T \cdot y \\ X_2^T \cdot y \end{bmatrix}_{K\times 1} &= \begin{bmatrix} \mathbf{1}^T \cdot [\mathbf{1}\hat{\beta}_1 + X_2\hat{\beta}_2] \\ X_2^T \cdot [\mathbf{1}\hat{\beta}_1 + X_2\hat{\beta}_2] \end{bmatrix} \end{aligned}$$

we will focus only on the first row, on the LHS we have $\mathbf{1}^T \cdot y = n.\bar{y}$, where $\bar{y}$ is the arithmetic mean, i.e $\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$. Similarly, $\mathbf{1}^T \mathbf{1} \hat{\beta}_1 = n\hat{\beta}_1$, and we define $\frac{\mathbf{1}^T X_2}{n} = \bar{\mathbf{x}}^T_{2_{1\times(k-1)}}$, so we obtain

$$\begin{aligned} n.\bar{y} &= n\hat{\beta}_1 + \bar{\mathbf{x}}_2^T \hat{\beta}_2 \\ \hat{\beta}_1 &= \bar{y} - \bar{\mathbf{x}}_2^T \hat{\beta}_2 \end{aligned}$$

that is the formula for the intercept estimate. Notice that, first we have to find the slope estimates $\hat{\beta}_2$, then we can obtain using this formula the intercept estimate.

3. **Sample Mean of Fitted Values**

Another implication of *including constant term* into the regression is that the sample mean of the fitted values is equal to the mean of the dependent variable. This exercise asks for the proof of this statement, namely we need to show that

$$\bar{y} = \bar{\hat{y}}$$

Proof. We obtain the fitted values by

$$\begin{aligned}
\hat{y} &= W\widehat{\beta} \\
W &= \begin{bmatrix} \mathbf{1} & X \end{bmatrix} \\
\widehat{y}_i &= \widehat{\beta}_1 + x_i^T \widehat{\beta}_2
\end{aligned}$$

where we split the $\widehat{\beta}$ vector as above, using the last formula we found in the previous example

$$\begin{aligned}
\widehat{y}_i &= \widehat{\beta}_1 + x_i^T \widehat{\beta}_2 = \bar{y} - \bar{\mathbf{x}}_2^T \hat{\beta}_2 + x_i^T \widehat{\beta}_2 \\
\sum_{i=1}^n \widehat{y}_i &= \sum_{i=1}^n (\bar{y} - \bar{\mathbf{x}}_2^T \hat{\beta}_2 + x_i^T \widehat{\beta}_2) = n\bar{y} - n\bar{\mathbf{x}}_2^T \hat{\beta}_2 + \sum_{i=1}^n x_i^T \widehat{\beta}_2 = \\
n\overline{\widehat{y}} &= n\bar{y} - n\bar{\mathbf{x}}_2^T \hat{\beta}_2 + n\bar{\mathbf{x}}_2^T \hat{\beta}_2 \\
\overline{\widehat{y}} &= \bar{y}.
\end{aligned}$$

4. **Decomposition of Sample Deviation (centered $\mathbf{R}^2$=coeffient of determination)**

We need to show (1.2.17, Hayashi) that

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\widehat{y}_i - \bar{y})^2 + \sum_{i=1}^n \widehat{\varepsilon}_i$$

Proof. First note the difference between the uncentered and centered R^2. The uncentered R^2 is derived from

$$\begin{aligned}
y^T y &= \widehat{y}^T \widehat{y} + \widehat{\varepsilon}^T \widehat{\varepsilon} \\
R_{uc}^2 &= 1 - \frac{\widehat{\varepsilon}^T \widehat{\varepsilon}}{y^T y} = \frac{\widehat{y}^T \widehat{y}}{y^T y}
\end{aligned}$$

given that we have a constant regressor along with other nonconstant regressors and recalling from the previos exercise $\overline{\widehat{y}} = \bar{y}$,

$$y^T y - n\bar{y}^2 = \widehat{y}^T \widehat{y} - n\overline{\widehat{y}}^2 + \widehat{\varepsilon}^T \widehat{\varepsilon}$$

first analyze the LHS

$$\begin{aligned}
y^T y - n\bar{y}^2 &= \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - 2\bar{y} \sum_{i=1}^n y_i + n\bar{y}^2 = \\
&= \sum_{i=1}^n y_i^2 - n\bar{y}^2 = y^T y - n\bar{y}^2
\end{aligned}$$

now we have to show that

$$\begin{aligned}
\sum_{i=1}^{n}(\widehat{y}_i - \bar{y})^2 &= \widehat{y}^T\widehat{y} - n\bar{\widehat{y}}^2 \\
\text{since } \bar{\widehat{y}} &= \bar{y} \\
\sum_{i=1}^{n}(\widehat{y}_i - \bar{y})^2 &= \sum_{i=1}^{n}(\widehat{y}_i - \bar{\widehat{y}})^2 = \widehat{y}^T\widehat{y} - n\bar{\widehat{y}}^2.
\end{aligned}$$

Hence we have shown that

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\widehat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}\widehat{\varepsilon}_i$$

From this the coefficient of determination R^2 is defined

$$R_c^2 = 1 - \frac{\widehat{\varepsilon}^T\widehat{\varepsilon}}{y^Ty - n\bar{y}^2} = \frac{\widehat{y}^T\widehat{y} - n\bar{\widehat{y}}^2}{y^Ty - n\bar{y}^2}$$

Provided that the regressors include a constant, $0 \leq R_c^2 \leq 1$, and it is a measure of the explanatory power of the nonconstant independent variables, in other words, numbers close to 1 tells us, most of the variation of the dependent variable can be explained with the variation of the independent variables included in the model.

5. **Some Sample Identities**

Two important identities to note are the **projection matrix P** and the **annihilator matrix M,** defined as follows

$$\begin{aligned}
P_{(n\times n)} &= X(X^TX)^{-1}X^T \\
M_{(n\times n)} &= \mathbf{I}_n - P = \mathbf{I}_n - X(X^TX)^{-1}X^T
\end{aligned}$$

In this exercise we will prove the nice properies of these two matrices.
Show that both are *symmetric* ($A^T = A$) and *idempotent* ($A = A^2$) matrices
Proof.

$$\begin{aligned}
P^T &= (X(X^TX)^{-1}X^T)^T = X(X^TX)^{-1}X^T = P \\
M^T &= \mathbf{I}_n^T - P^T = M \\
PP &= X(X^TX)^{-1}X^TX(X^TX)^{-1}X^T = X(X^TX)^{-1}X^T = P \\
MM &= (\mathbf{I}_n - P)\ (\mathbf{I}_n - P) = \mathbf{I}_n - P - P + PP = \mathbf{I}_n - P = M
\end{aligned}$$

Show that

$$\begin{aligned}
PX &= X \Rightarrow \textit{projection} \\
MX &= 0 \Rightarrow \textit{annihilator}
\end{aligned}$$

Proof.

$$\begin{aligned} PX &= X(X^TX)^{-1}X^TX = X \\ MX &= (\mathbf{I}_n - P)X = X - X = 0. \end{aligned}$$

Show that

$$\begin{aligned} \widehat{y} &= Py \\ \widehat{\varepsilon} &= My = M\varepsilon \end{aligned}$$

Proof.

$$\begin{aligned} \widehat{y} &= X\widehat{\beta} = X(X^TX)^{-1}X^Ty = Py \\ \widehat{\varepsilon} &= y - X\widehat{\beta} = \mathbf{I}_n y - Py = (\mathbf{I}_n - P)y = My \\ \widehat{\varepsilon} &= My = M(X\beta + \varepsilon) = MX\beta + M\varepsilon = M\varepsilon \\ \text{since } MX &= 0. \end{aligned}$$

these matrices provide useful shortcuts for calculations, e.g

$$\begin{aligned} SSR &= \varepsilon^T M\varepsilon \\ \widehat{\varepsilon} &= M\varepsilon \\ \widehat{\varepsilon}^T\widehat{\varepsilon} &= \varepsilon^T MM\varepsilon = \varepsilon^T M\varepsilon. \\ SSR &= (y - X\widehat{\beta})^T(y - X\widehat{\beta}) = \widehat{\varepsilon}^T\widehat{\varepsilon} = \varepsilon^T M\varepsilon \\ \widehat{\varepsilon}^T\widehat{\varepsilon} &= y^T MMy = y^T My = SSR. \end{aligned}$$

6. **Effects of Change in Unit of Measurement**

We might be interested to see how the changes in measurement units affect certain values, such as

i.) OLS estimate

ii.) SSR

iii.) R^2

Below we analyze three different cases and compare with the benchmark model in order to see the effects of measurement unit change on abovementioned values;

$$\begin{aligned} \text{Benchmark model} &: \quad y = W\beta + \varepsilon \\ \widehat{\beta} &= (X^TX)^{-1}X^Ty \\ SSR &: \quad \widehat{\varepsilon}^T\widehat{\varepsilon} \end{aligned}$$

a.) Change in dependent variable:

$$\frac{1}{a}y = W\gamma + \eta$$

i.) OLS estimate: $\widehat{\gamma} = (W^T W)^{-1} W^T \frac{y}{a} = \frac{\widehat{\beta}}{a}$.

ii.) SSR: $\widehat{\eta} = \frac{y}{a} - W\widehat{\gamma} = \frac{y}{a} - W\frac{\widehat{\beta}}{a} = \frac{\widehat{\varepsilon}}{a} \Rightarrow \widehat{\eta}^T\widehat{\eta} = \frac{\widehat{\varepsilon}^T\widehat{\varepsilon}}{a^2} \Rightarrow SSR_\eta = \frac{1}{a^2}SSR_\varepsilon$.

iii.) $\mathbf{R}^2$:$\mathbf{R}^2_\eta = 1 - \frac{SSR_\eta}{TSS/a^2} = 1 - \frac{\frac{1}{a^2}SSR_\varepsilon}{\frac{1}{a^2}TSS} = \mathbf{R}^2_\varepsilon$.

b.) Change in independent variables:

$$y = \frac{1}{a}W\gamma + \eta$$

i.) OLS estimate: $\widehat{\gamma} = (\frac{W^T}{a}\frac{W}{a})^{-1}\frac{W^T}{a}y = a\widehat{\beta}$.

ii.) SSR: $\widehat{\eta} = y - \frac{W}{a}\widehat{\gamma} = y - \frac{W}{a}a\widehat{\beta} = \widehat{\varepsilon} \Rightarrow SSR_\eta = SSR_\varepsilon$.

iii.) $\mathbf{R}^2$: $TSS_\eta = TSS_e \Rightarrow \mathbf{R}^2_\eta = \mathbf{R}^2_\varepsilon$.

c.) Change in both dependent and independent variables:

$$\frac{1}{a}y = \frac{1}{a}W\gamma + \eta$$

i.) OLS estimate: $\widehat{\gamma} = (\frac{W^T}{a}\frac{W}{a})^{-1}\frac{W^T}{a}\frac{1}{a}y = \widehat{\beta}$.

ii.) SSR: $\widehat{\eta} = \frac{1}{a}y - \frac{W}{a}\widehat{\gamma} = \frac{1}{a}y - \frac{W}{a}\widehat{\beta} = \frac{1}{a}\widehat{\varepsilon} \Rightarrow SSR_\eta = \frac{1}{a^2}SSR_\varepsilon$.

iii.) $\mathbf{R}^2$: $TSS_\eta = \frac{1}{a^2}TSS_e \Rightarrow \mathbf{R}^2_\eta = \mathbf{R}^2_\varepsilon$.

Note that, in all three cases $\mathbf{R}^2$ remain the same, so one cannot change the fit of the model by simply changing the unit of measurement of the variables.