INTRODUCTORY ECONOMETRICS

Lesson 2b

Dr Javier Fernández

etpfemaj@ehu.es

Dpt. of Econometrics & Statistics

UPV—EHU

2.3b OLS in the GLRM.

GLRM: the PRF

- Recall: model with K explanatory variables:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_K X_{Kt} + u_t, \qquad (2)$$
$$Y = X\beta + u$$

  is called GLRM.
- Population Regression Function (PRF):
  $E(u) = 0 \rightsquigarrow$ *systematic part* or PRF:

$$E(Y_t) = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_K X_{Kt}$$
$$E(Y) = X\beta$$

- Interpretation of the coefficients:
  - $\beta_0 = E(Y_t | X_{1t} = X_{2t} = \cdots = X_{Kt} = 0)$: Expected value of Y_t when all explanatory variables are equal to zero.
  - $\beta_k = \frac{\partial E(Y_t)}{\partial X_{kt}} \simeq \frac{\Delta E(Y_t)}{\Delta X_{kt}}, \quad k = 1 \dots K$: Increase in (expected) value Y_t when $X_k \uparrow$ one unit (c.p.).

The Sample Regression Function (SRF)

- Objective of GLRM: To obtain estimator $\widehat{\beta} = (\widehat{\beta}_0, \widehat{\beta}_1 \dots, \widehat{\beta}_K)'$ of unknown parameter vector in (2).
  $\widehat{\beta} \rightsquigarrow$ estimated model, fit or SRF:

$$\widehat{Y}_t = \widehat{\beta}_0 + \widehat{\beta}_1 X_{1t} + \cdots + \widehat{\beta}_K X_{Kt}$$
$$\widehat{Y} = X\widehat{\beta}$$

- Notes:
  - Disturbances in PRF:

$$u_t = Y_t - E(Y_t) = Y_t - \beta_0 - \beta_1 X_{1t} - \cdots - \beta_K X_{Kt}$$
$$u = Y - E(Y) = Y - X\beta$$

  - Residuals in SRF:

$$\widehat{u}_t = Y_t - \widehat{Y}_t = Y_t - \widehat{\beta}_0 - \widehat{\beta}_1 X_{1t} - \cdots - \widehat{\beta}_K X_{Kt}$$
$$\widehat{u} = Y - \widehat{Y} = Y - X\widehat{\beta}$$

- Residuals are to the SRF what disturbances are to the PRF.

Estimation: OLS

- apply Least-Squares fit to GLRM: $Y = X\beta + u$,
- either in observation form:

$$\min_{\beta_0 \dots \beta_K} \sum_{t=1}^{T} u_t^2 \text{ where } u_t = Y_t - \beta_0 - \beta_1 X_{1t} - \cdots - \beta_K X_{Kt}$$

- or in matrix form:
  [recall:

$$u' = (u_1, u_2, \dots, u_T) \qquad u = \begin{pmatrix} u_1 \\ u_2 \\ \dots \\ u_T \end{pmatrix}$$

  so $u'u = u_1^2 + u_2^2 + \cdots + u_T^2 = \sum_{t=1}^{T} u_t^2$]
- that is

$$\min_{\beta} u'u \quad \text{where} \quad u = Y - X\beta$$

Note: vector derivatives

- Let $u = u(\beta)$: derivs of $c\,u$ and $c\,u^2$ with respect to β:

$$\frac{\partial}{\partial\beta}(c\,u) = c\,\frac{\partial u}{\partial\beta} \qquad \text{and} \qquad \frac{\partial}{\partial\beta}u^2 = 2\,u\,\frac{\partial u}{\partial\beta}$$

- With vectors and matrices this is quite similar:

- The derivative of the linear combination $u'c$

$$\underset{(1\times n)}{u'}\;\underset{(n\times 1)}{c} \qquad \left(=\textstyle\sum_{i=1}^{n} c_i u_i,\ \textit{i.e.}\ \text{scalar!!}\right)$$

with respect to $\underset{(k\times 1)}{\beta}$ is: $\frac{\partial(u'c)}{\partial\beta} = \frac{\partial u'}{\partial\beta}c$

- The derivative of the sum of squares $u'u$

$$\underset{(1\times n)}{u'}\;\underset{(n\times 1)}{u} \qquad \left(=\textstyle\sum_{i=1}^{n} u_i^2,\ \textit{i.e.}\ \text{scalar!!}\right)$$

with respect to $\underset{(k\times 1)}{\beta}$ is: $\frac{\partial(u'u)}{\partial\beta} = 2\,\frac{\partial u'}{\partial\beta}u$

1st.o.c. in matrix form

$$\min_{\beta}(u'u) \quad \text{where} \quad u = Y - X\beta$$

First derivatives of SS $u'u$ with respect to β:

$$\begin{aligned}\frac{\partial u'u}{\partial\beta} &= 2\,\frac{\partial u'}{\partial\beta}u \\ &= 2\,\frac{\partial(Y' - \beta'X')}{\partial\beta}u \\ &= -2\,X'u\end{aligned}$$

in the minimum:

$$\text{1st.o.c.:} \quad \underset{(K+1\times T)}{X'}\;\underset{(T\times 1)}{\widehat{u}} = 0_{K+1}$$

Estimation: Normal equations & LSE of β

Solving the 1st.o.c. we obtain the normal equations: $X'(Y - X\widehat{\beta}) = 0 \Rightarrow$

$$\underset{(K+1\times 1)}{X'Y} = \underset{(K+1\times K+1)}{X'X}\;\underset{(K+1\times 1)}{\widehat{\beta}} \qquad (3)$$

Whence premultiplying by $(X'X)^{-1}$ we obtain the OLS estimator:

$$\widehat{\beta}_{\text{OLS}} = (X'X)^{-1}X'Y$$

Estimation: LSE of β (cont)

- where $X'X$ is a $[K{+}1 \times K{+}1]$ matrix: [recall X & Y? $\longrightarrow$]
-

$$\underset{(K+1\times K+1)}{X'X} = \begin{bmatrix} T & \sum X_{1t} & \sum X_{2t} & \dots & \sum X_{Kt} \\ \sum X_{1t} & \sum X_{1t}^2 & \sum X_{1t}X_{2t} & \dots & \sum X_{1t}X_{Kt} \\ \dots & \dots & \dots & \dots & \dots \\ \sum X_{Kt} & \sum X_{Kt}X_{1t} & \sum X_{Kt}X_{2t} & \dots & \sum X_{Kt}^2 \end{bmatrix}$$

- and $X'Y$ and $\widehat{\beta}$ are $[K{+}1 \times 1]$ vectors:

$$\underset{(K+1\times 1)}{X'Y} = \begin{bmatrix} \sum Y_t \\ \sum X_{1t}Y_t \\ \dots \\ \sum X_{Kt}Y_t \end{bmatrix} \qquad \underset{(K+1\times 1)}{\widehat{\beta}} = \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \dots \\ \widehat{\beta}_K \end{bmatrix}$$

OLS estimator with centred (demeaned) data

An alternative way to obtain the OLS estimator is

$$\widehat{\beta}^{\star}_{\text{OLS}} = (x'x)^{-1}x'y$$

for the model coefficients.

…together with the estimated intercept obtained from the first normal equation

$$\widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1\overline{X}_1 - \cdots - \widehat{\beta}_K\overline{X}_K$$

Note: special case with $K = 1 \rightsquigarrow$ identical formulae as in SLRM!! (Prove it!!)

2.4b Properties of the SRF.

Properties of residuals and SRF (1)

$$\left.\begin{array}{c}\widehat{\beta} \\ \widehat{\beta}^{\star} \rightsquigarrow \widehat{\beta}_0\end{array}\right\} \rightsquigarrow \widehat{Y} = X\widehat{\beta} \rightsquigarrow \widehat{u} = Y - \widehat{Y}$$

1. residuals add up to zero: $\sum \widehat{u}_t = 0$
   Demo: directly from 1st.o.c.:

$$X'\widehat{u} = 0 \Rightarrow \begin{bmatrix} \sum_1^T \widehat{u}_t \\ \sum_1^T X_{1t}\widehat{u}_t \\ \sum_1^T X_{2t}\widehat{u}_t \\ \cdots \\ \sum_1^T X_{Kt}\widehat{u}_t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix}$$

□

2. $\overline{\widehat{Y}} = \overline{Y}$
3. the SRF passes thru vector $(\overline{X}_1, \ldots \overline{X}_K, \overline{Y})$: $\overline{Y} = \widehat{\beta}_0 + \widehat{\beta}_1\overline{X}_1 + \cdots + \widehat{\beta}_K\overline{X}_K$

   Note: These properties 1 thru 3 are fulfilled if the regression has an intercept; that is, if X has a column of "ones".

Properties of residuals and SRF (2)

4. residuals orthogonal to explanatory v.: $X'\widehat{u} = 0$
   Demo: directly from 1st.o.c. (see 1) or, alternatively:

$$\begin{aligned} X'\widehat{u} &= X'(Y - X\widehat{\beta}) = X'Y - X'X\widehat{\beta} \\ &= X'Y - \underbrace{X'X(X'X)^{-1}}_{=I_{K+1}}X'Y = 0 \end{aligned}$$

□

5. residuals orthogonal to explained part of Y: $\widehat{Y}'\widehat{u} = 0$
   Demo: $\widehat{Y}'\widehat{u} = (X\widehat{\beta})'\widehat{u} = \widehat{\beta}'\underbrace{X'\widehat{u}}_{=0} = 0$ □

2.5b Goodness of Fit: Coefficient of Determination (R^2) & Estimation of the Error Variance.

Goodness of fit: R^2 Revisited

Recall (same as before but now we'll do it with vectors):

$$Y'Y = (\widehat{Y}' + \widehat{u}')(\widehat{Y} + \widehat{u})$$
$$= \widehat{Y}'\widehat{Y} + \widehat{u}'\widehat{u} + 2\widehat{Y}'\widehat{u}$$
$$= \widehat{Y}'\widehat{Y} + \widehat{u}'\widehat{u} \quad \text{(from prop 5)}$$

$$Y'Y - T\overline{Y}^2 = \widehat{Y}'\widehat{Y} - T\overline{\widehat{Y}}^2 + \widehat{u}'\widehat{u} \qquad \text{(from prop 2)}$$

$$\underset{(TSS)}{y'y} = \underset{(ESS)}{\hat{y}'\hat{y}} + \underset{(RSS)}{u'u}$$

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$
$$0 \leq R^2 \leq 1$$

Goodness of fit: R^2 Revisited (cont)

Note 1: R^2 measures the proportion of the dependent variable variation explained by the variation of (a linear combination of) the explanatory variables.

Note 2:

$$\text{no intercept} \Rightarrow \begin{cases} \nexists \text{1st row of 1st.o.c.} \rightsquigarrow \begin{cases} \sum \widehat{u}_t \neq 0, \\ \overline{\widehat{Y}} \neq \overline{Y}, \end{cases} \\ \text{not valid} R^2 \text{ (Remember!)} \end{cases}$$

Estimation of $\mathrm{Var}(u_t)$

$$\sigma^2 = \mathrm{Var}(u_t) = \mathrm{E}(u_t^2) \simeq \frac{1}{T}\sum_{t=1}^{T} u_t^2$$

but with residuals, they must satisfy $K+1$ linear relationships in $X'\widehat{u} = 0$ so we loose $K+1$ degrees of freedom:

$$\widehat{\sigma}^2 = \frac{1}{T-K-1}\sum_{t=1}^{T}\widehat{u}_t^2$$

Therefore we propose the following estimator:

$$\widehat{\sigma}^2 = \frac{\mathrm{RSS}}{T-K-1}$$

which clearly is an unbiased estimator:

Demo:

$$\mathrm{E}(\widehat{\sigma}^2) = \frac{\mathrm{E}(\mathrm{RSS})\,(^*)}{T-K-1} = \frac{T-K-1}{T-K-1} = \sigma^2$$

□ (* see textbook)

2.6 Finite-sample Properties of the Least-Squares Estimator. The Gauss-Markov Theorem.

Properties of the OLS Estimator (1)

The estimator $\widehat{\beta}_{\text{OLS}} = (X'X)^{-1}X'Y$ has the following properties:

- **Linear:** $\widehat{\beta}_{\text{OLS}}$ is a linear combination of disturbances:

$$\begin{aligned}\widehat{\beta} &= (X'X)^{-1}X'(X\beta + u)\\ &= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u\\ &= \beta + (X'X)^{-1}X'u\\ &= \beta + \Gamma' u\end{aligned}$$

- **Unbiased:** Since $\mathsf{E}(u) = 0$, $\widehat{\beta}_{\text{OLS}}$ is unbiased:

$$\begin{aligned}\mathsf{E}(\widehat{\beta}) &= \mathsf{E}(\beta + \Gamma' u)\\ &= \beta + \Gamma'\mathsf{E}(u)\\ &= \beta\end{aligned}$$

Properties of the OLS Estimator (2)

- **Variance:** Recall:

$$\begin{aligned}\text{Var}(u) &= \sigma^2 I_T,\\ \widehat{\beta} &= \beta + (X'X)^{-1}X'u,\end{aligned}$$

$$\begin{aligned}\text{Var}(\widehat{\beta}) &= \mathsf{E}\left((\widehat{\beta} - \beta)(\widehat{\beta} - \beta)'\right)\\ &= \mathsf{E}\left((X'X)^{-1}X'uu'X(X'X)^{-1}\right)\\ &= (X'X)^{-1}X'\,\mathsf{E}(uu')\,X(X'X)^{-1}\\ &= (X'X)^{-1}X'\,\sigma^2 I_T\,X(X'X)^{-1}\\ &= \sigma^2(X'X)^{-1}X'X(X'X)^{-1}\end{aligned}$$

$$\boxed{\text{Var}(\widehat{\beta}) = \sigma^2(X'X)^{-1}}$$

Properties of the OLS Estimator (2cont)

$$\text{Var}(\widehat{\beta}) = \begin{bmatrix} \text{Var}(\widehat{\beta}_0) & \text{Cov}(\widehat{\beta}_0, \widehat{\beta}_1) & \ldots & \text{Cov}(\widehat{\beta}_0, \widehat{\beta}_K)\\ \text{Cov}(\widehat{\beta}_1, \widehat{\beta}_0) & \text{Var}(\widehat{\beta}_1) & \ldots & \text{Cov}(\widehat{\beta}_1, \widehat{\beta}_K)\\ \ldots & \ldots & \ldots & \ldots\\ \text{Cov}(\widehat{\beta}_K, \widehat{\beta}_0) & \text{Cov}(\widehat{\beta}_K, \widehat{\beta}_1) & \ldots & \text{Var}(\widehat{\beta}_K)\end{bmatrix}$$

$$\sigma^2(X'X)^{-1} = \sigma^2\begin{bmatrix} a_{00} & a_{00} & a_{01} & \ldots & a_{0K}\\ a_{10} & a_{11} & a_{12} & \ldots & a_{1K}\\ a_{20} & a_{21} & a_{22} & \ldots & a_{2K}\\ \ldots & \ldots & \ldots & \ldots & \ldots\\ a_{K0} & a_{K1} & a_{K2} & \ldots & a_{KK}\end{bmatrix}$$

i.e. a_{kk} is the $(k+1, k+1)$-element of matrix $(X'X)^{-1}$:

$$\begin{aligned}\text{Var}(\widehat{\beta}_k) &= \sigma^2 a_{kk}\\ \text{Cov}(\widehat{\beta}_k, \widehat{\beta}_i) &= \sigma^2 a_{ki}\end{aligned}$$

The Gauss-Markov Theorem

"Given the basic assumptions of GLRM, the OLS estimator is that of minimum variance (best) among all the linear and unbiased estimators"

$$\widehat{\beta}_{\text{OLS}} = \textbf{BLUE} = \textbf{B}_{\text{est}}\textbf{L}_{\text{inear}}\textbf{U}_{\text{nbiased}}\textbf{E}_{\text{stimator}}$$

Demo:

Let $\widetilde{\beta}$ be some other linear and unbiased estimator:

$$\widetilde{\beta} = D'Y = D'(X\beta + u) = D'X\beta + D'u$$

$$\mathrm{E}(\widetilde{\beta}) = D'X\beta + D'\mathrm{E}(u) = D'X\beta = \beta \Rightarrow \boxed{D'X = I_K}$$

then $\widetilde{\beta} = \beta + D'u \quad \rightsquigarrow \quad \widetilde{\beta} - \beta = D'u$
and its variance:

$$\mathrm{Var}(\widetilde{\beta}) = E\left[(\widetilde{\beta} - \beta)(\widetilde{\beta} - \beta)'\right] = \mathrm{E}(D'uu'D)$$

$$= D'\,\mathrm{E}(uu')\,D = D'\,\sigma^2 I_T\,D = \sigma^2 D'D$$

The Gauss-Markov Theorem (cont)

. . . The difference between both covariance matrices is a positive definite matrix:

$$\begin{aligned}
\mathrm{Var}(\widetilde{\beta}) - \mathrm{Var}(\widehat{\beta}) &= \sigma^2 D'D - \sigma^2 (X'X)^{-1} \\
&= \sigma^2 \left[D'D - (X'X)^{-1}\right] \\
&= \sigma^2 \left[D'D - D'X(X'X)^{-1}X'D\right] \\
&= \sigma^2 D' \underbrace{\left[I_T - X(X'X)^{-1}X'\right]}_{M} D \\
&= \sigma^2 D'(MM)D \\
&= \sigma^2 (D'M)(M'D) = D^{\star\prime}D^{\star} \\
&> 0
\end{aligned}$$

I.e. in particular **all** individual variances will be bigger than their OLS counterpart.

2.3c OLS: Useful expressions & Timeline.

Useful expressions for SS

- $$TSS = \sum (Y_t - \overline{Y})^2 = \sum Y_t^2 - T\overline{Y}^2 = Y'Y - T\overline{Y}^2$$

- $$ESS = \sum (\widehat{Y}_t - \overline{\widehat{Y}})^2 = \sum \widehat{Y}_t^2 - T\overline{\widehat{Y}}^2 = \sum \widehat{Y}_t^2 - T\overline{Y}^2 = \widehat{Y}'\widehat{Y} - T\overline{Y}^2$$
  $$= (X\widehat{\beta})'(X\widehat{\beta}) - T\overline{Y}^2 = \widehat{\beta}'\underbrace{X'X\widehat{\beta}}_{X'Y} - T\overline{Y}^2 = \widehat{\beta}'X'Y - T\overline{Y}^2$$

- $$RSS = \sum \widehat{u}_t^2 = \widehat{u}'\widehat{u} \qquad = \sum Y_t^2 - \sum \widehat{Y}_t^2 = Y'Y - \widehat{\beta}'X'Y$$

Main expressions & Timeline

- $Y = X\beta + u$
- $(X'X)^{-1}$ $\quad X'Y$
- $\widehat{\beta} = (X'X)^{-1}X'Y$
- $ESS = \widehat{\beta}'X'Y - T\overline{Y}^2$ (needs $\overline{Y}$!)
- $TSS = Y'Y - T\overline{Y}^2$
- $RSS = Y'Y - \widehat{\beta}'X'Y$ (no $\overline{Y}$!)
- $R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$
- $\widehat{\sigma}^2 = \frac{RSS}{T-K-1}$
- $\widehat{\text{Var}(\widehat{\beta})} = \widehat{\sigma}^2 (X'X)^{-1}$