INTRODUCTORY ECONOMETRICS

3rd year LE & LADE

LESSON 2

Dr Javier Fernández-Macho

etpfemaj@ehu.es

Dpt. of Econometrics & Statistics

UPV—EHU

2 The Linear Regression Model (I). Specification and Estimation.

2.1 Specification of the General Linear Regression Model (GLRM).

Specification of the GLRM (1)

- **Objective:** Quantifying the relationship between:
  - a variable Y and
  - a set of K explanatory variables $X_1, X_2, \dots, X_K$,
  - by means of a linear model.
- **Starting point:**
  - a linear model: $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_K X_K + u$,
  - a data sample of size T: $Y_t, X_{1t}, X_{2t}, \dots, X_{Kt},\ t = 1 \dots T$, where

$$Y_t = t\text{-th obs of } Y,$$

$$X_{kt} = t\text{-th obs of } X_k,\ k = 1, 2 \dots K.$$

Specification of the GLRM (2)

- GLRM:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_K X_{Kt} + u_t,\ t = 1, 2 \dots T,$$

  whose elements are (recall):
  - Y: dependent variable,
  - $X_k,\ k = 1 \dots K$: explanatory variables,
  - β_0: intercept, $\beta_k,\ k = 1 \dots K$: coefficients (parameters to be estimated),
  - u: (non-observable random) error or disturbance, that allows for:
    - variables not included in the model,
    - random behaviour of economic agents,
    - measurement errors.

The GLRM in observation form

the model

$$Y_t = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_K X_{Kt} + u_t, \; t = 1, 2 \ldots T,$$

implies for each observation:

$$Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} + \cdots + \beta_K X_{K1} + u_1$$
$$Y_2 = \beta_0 + \beta_1 X_{12} + \beta_2 X_{22} + \cdots + \beta_K X_{K2} + u_2$$
$$\ldots\ldots$$
$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_K X_{Kt} + u_t$$
$$\ldots\ldots$$
$$Y_T = \beta_0 + \beta_1 X_{1T} + \beta_2 X_{2T} + \cdots + \beta_K X_{KT} + u_T$$

The GLRM in matrix form (1)

or else in matrix form:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \ldots \\ Y_t \\ \ldots \\ Y_T \end{bmatrix} = \begin{bmatrix} \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} + \cdots + \beta_K X_{K1} \\ \beta_0 + \beta_1 X_{12} + \beta_2 X_{22} + \cdots + \beta_K X_{K2} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_K X_{Kt} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \beta_0 + \beta_1 X_{1T} + \beta_2 X_{2T} + \cdots + \beta_K X_{KT} \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \ldots \\ u_t \\ \ldots \\ u_T \end{bmatrix}$$

The GLRM in matrix form (2)

- that is:

$$\underset{(T \times 1)}{\underset{Y}{\begin{bmatrix} Y_1 \\ Y_2 \\ \ldots \\ Y_t \\ \ldots \\ Y_T \end{bmatrix}}} = \underset{(T \times K{+}1)}{\underset{X}{\begin{bmatrix} 1 & X_{11} & X_{21} & \ldots & X_{K1} \\ 1 & X_{12} & X_{22} & \ldots & X_{K2} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & X_{1t} & X_{2t} & \ldots & X_{Kt} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & X_{1T} & X_{2T} & \ldots & X_{KT} \end{bmatrix}}} \underset{(K{+}1 \times 1)}{\underset{\beta}{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix}}} + \underset{(T \times 1)}{\underset{u}{\begin{bmatrix} u_1 \\ u_2 \\ \ldots \\ u_t \\ \ldots \\ u_T \end{bmatrix}}}$$

- $$\boxed{Y = X\beta + u.}$$

↰

2.2 Basic (Classical) Assumptions. Interpretation.

Basic Assumptions of the GLRM (1)

1. Assumptions about the relationship:
   - Model is correctly specified:

$$X_k \text{ explains } Y \Leftrightarrow X_k \in \text{model.}$$

2. Assumptions about the parameters:
   - they are constant throughout the sample,
   - they appear linearly (*i.e.* a constant plus coefficients)
     - $Y_t = \beta_0 + \beta_1 X_t + u_t$
   - Note: but vars $Y, X_1, X_2, \ldots$ may be transformations:
     - $Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_t^2 + \beta_3 \dfrac{1}{X_t} + u_t$
     - $Y_t = A X_{1t}^{\beta_1} X_{2t}^{\beta_2} \mathrm{e}^{u_t}$ (Why?)
     - and this? $\quad Y_t = \beta_0 + \beta_1 \dfrac{1}{X_t - \beta_2} + u_t$
     - and these other?
       $\ln Y_t = \beta_0 X_t^{\beta_1} u_t; \qquad Y_t = \beta_0 X_t^{\beta_1} + u_t$
       $Y_t = \beta_1 X_{1t} + \beta_2 X_{1t} X_{2t} + u_t; \quad Y_t = \beta_0 + \beta_1 X_{1t}^{X_{2t}} + u_t$

Basic Assumptions of the GLRM (2)

3. Assumptions about the explanatory variables:
   (a) $X_1, \ldots, X_K$, are quantitative and fixed (*i.e.* not random).
   (b) $X_1, \ldots, X_K$, are linearly independent:
   - $\nexists X_k | X_k =$ lin. comb. of others (Why?)

- Examples of not valid cases:
  - $Y_t = \beta_0 + \beta_1 X_t + \beta_2 (2X_t + 3) + u_t$
  - $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 (X_{1t} + X_{2t}) + u_t$
- Examples of valid cases:
  - $Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_t^2 + u_t$
  - $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{1t} X_{2t} + u_t$

Basic Assumptions of the GLRM (3)

4. Assumptions about the disturbance term:
   (a) Zero mean:
   $\mathsf{E}(u_t) = 0 \quad \forall t$ (isn't obvious?).
   (b) Homoscedastic:
   $\mathsf{Var}(u_t) = \mathsf{E}(u_t^2) = \sigma_u^2 \; (= \sigma^2)$ const ($\forall t$).
   (c) Serially uncorrelated:
   $\mathsf{Cov}(u_t, u_s) = \mathsf{E}(u_t u_s) = 0 \quad \forall t \neq s.$
   (d) Normally distributed$^{(*)}$:
   $u_t \sim \mathcal{N} \quad \forall t.$ (* added)

- Assumptions 4a–4d jointly:

$$u_t \sim \text{iid}\, \mathcal{N}(0, \sigma_u^2)$$

Basic Assumptions in matrix form (1)

- from 4a: Mean Vector:

$$\underset{(T \times 1)}{\mathsf{E}(u)} = \begin{bmatrix} \mathsf{E}(u_1) \\ \mathsf{E}(u_2) \\ \vdots \\ \mathsf{E}(u_T) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = 0_T$$

- from 4b and 4c: Covariance Matrix:

$$\underset{(T \times T)}{\mathsf{E}(uu')} = \begin{bmatrix} \mathsf{E}(u_1^2) & \mathsf{E}(u_1 u_2) & \ldots & \mathsf{E}(u_1 u_T) \\ \mathsf{E}(u_2 u_1) & \mathsf{E}(u_2^2) & \ldots & \mathsf{E}(u_2 u_T) \\ \ldots & \ldots & \ldots & \ldots \\ \mathsf{E}(u_T u_1) & \mathsf{E}(u_T u_2) & \ldots & \mathsf{E}(u_T^2) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_u^2 & 0 & \ldots & 0 \\ 0 & \sigma_u^2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \sigma_u^2 \end{bmatrix} = \sigma_u^2 I_T$$

Basic Assumptions in matrix form (2)

- more compactly:

$$\underset{(T\times 1)}{u} \sim \big(\underset{(T\times 1)}{0} \;, \underset{(T\times T)}{\sigma_u^2 I_T} \big)$$

- plus 4d:

$$\underset{(T\times 1)}{u} \sim \mathcal{N}\big(\underset{(T\times 1)}{0} \;, \underset{(T\times T)}{\sigma_u^2 I_T} \big)$$

2.3a Ordinary Least Squares (OLS) in a Single Linear Regression Model (SLRM).

SLRM: the PRF

- With $K = 1 \quad \rightsquigarrow \quad Y_t = \beta_0 + \beta_1 X_{1t} + u_t,$

$$\text{(SLRM):} \quad Y_t = \alpha + \beta X_t + u_t. \qquad (1)$$

- Population Regression Function (PRF):
  $\mathsf{E}(u_t) = 0 \quad \rightsquigarrow$ *systematic part* or PRF:

$$\mathsf{E}(Y_t) = \alpha + \beta X_t$$

- Interpretation of the parameters:
  - $\alpha = \mathsf{E}(Y_t | X_t = 0)$: Expected value of Y_t
    when the explanatory variable is zero.
  - $\beta = \dfrac{\partial \mathsf{E}(Y_t)}{\partial X_t} \simeq \dfrac{\Delta \mathsf{E}(Y_t)}{\Delta X_t}$: Increase in (expected) value of Y_t
    when $X \uparrow$ one unit (*c.p.*).
- Objective: To obtain estimates $\widehat{\alpha}$, $\widehat{\beta}$
  of the unknown parameters α, β in (1).

The Sample Regression Function (SRF)

- $\widehat{\alpha}$, $\widehat{\beta} \quad \rightsquigarrow$ model estimate or SRF:

$$\widehat{Y}_t = \widehat{\alpha} + \widehat{\beta} X_t$$

- Interpretation of the estimates:
  - $\widehat{\alpha} = (\widehat{Y}_t | X_t = 0)$: Estimated value of Y_t
    when the explanatory variable is zero.
  - $\widehat{\beta} = \dfrac{\partial \widehat{Y}_t}{\partial X_t} \simeq \dfrac{\Delta \widehat{Y}_t}{\Delta X_t}$: Estimated increase in Y_t
    when $X \uparrow$ one unit (*c.p.*).
- Note difference: an estimator (a formula)
  vs. an estimate (a number).

Disturbances vs. Residuals

- Disturbances in PRF:

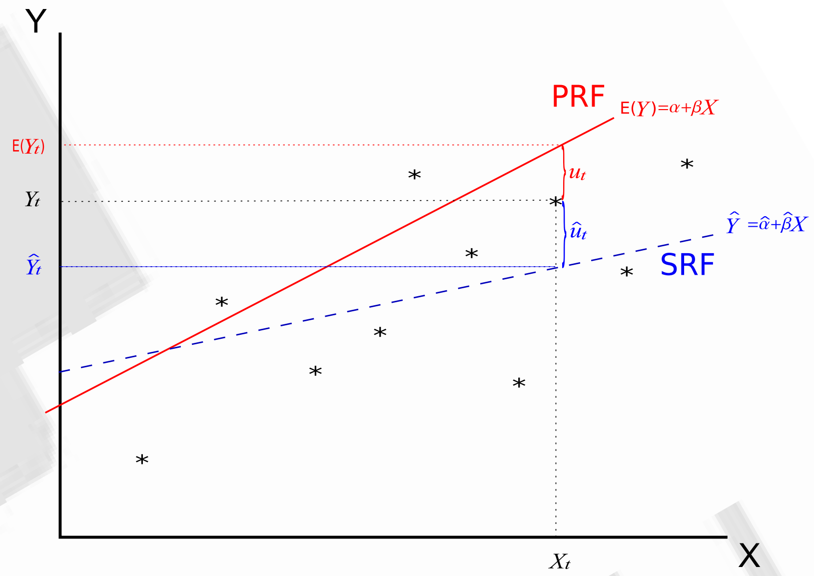
$$u_t = Y_t - \mathsf{E}(Y_t) = Y_t - \alpha - \beta X_t$$

- Residuals in SRF:

$$\widehat{u}_t = Y_t - \widehat{Y}_t = Y_t - \widehat{\alpha} - \widehat{\beta} X_t$$

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

SLRM: PRF and SRF

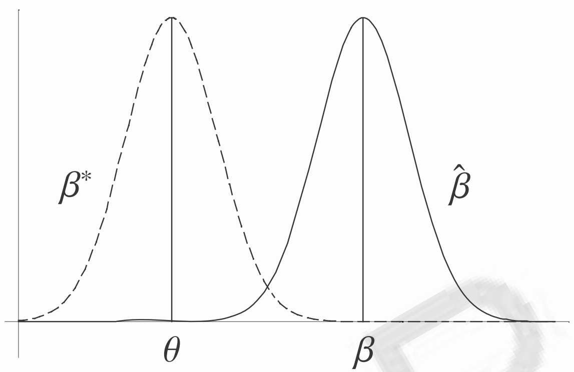


Estimation: Desired Properties (1)

Let $\widehat{\beta}$ be an estimator of β...

Unbiasedness:

$$\mathsf{E}(\widehat{\beta}) = \beta \quad \Leftrightarrow \quad \widehat{\beta} \text{ unbiased}$$

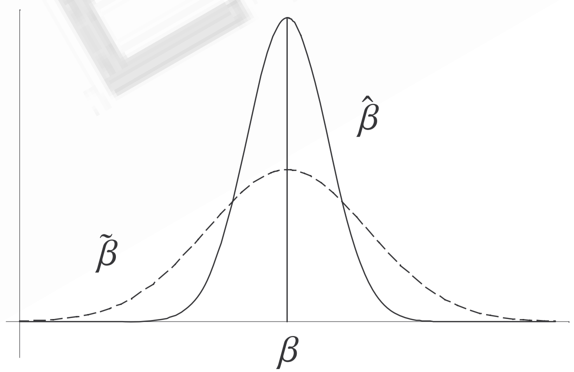


Estimation: Desired Properties (2)

Let $\widehat{\beta}$ and $\widetilde{\beta}$ be unbiased estimators of β...

Relative efficiency:

$$\mathsf{Var}(\widehat{\beta}) \leq \mathsf{Var}(\widetilde{\beta}) \quad \Leftrightarrow \quad \widehat{\beta} \text{ relatively efficient}$$



Estimation: OLS criteria

SLRM: $Y_t = \alpha + \beta X_t + u_t$,

- apply Least-Squares fit:

$$\min_{\alpha,\beta} \sum_{t=1}^{T} u_t^2 \quad \text{where} \quad u_t = Y_t - \alpha - \beta X_t :$$

- First derivatives:
  - $$\frac{\partial \sum u_t^2}{\partial \alpha} = 2\sum u_t \frac{\partial u_t}{\partial \alpha} = 2\sum u_t(-1)$$
  - $$\frac{\partial \sum u_t^2}{\partial \beta} = 2\sum u_t \frac{\partial u_t}{\partial \beta} = 2\sum u_t(-X_t)$$
- 1st.o.c. (minimum) $\Rightarrow$ first derivatives are zero:
  - $$\sum \widehat{u}_t = \sum (Y_t - \widehat{\alpha} - \widehat{\beta} X_t) = 0$$
  - $$\sum \widehat{u}_t X_t = \sum (Y_t X_t - \widehat{\alpha} X_t - \widehat{\beta} X_t^2) = 0$$

Estimation: Normal equations & LSE of α

- From the above 1st.o.c's:

$$\sum (Y_t - \widehat{\alpha} - \widehat{\beta} X_t) = 0$$
$$\sum (Y_t X_t - \widehat{\alpha} X_t - \widehat{\beta} X_t^2) = 0$$

- we obtain the Normal Equations:

$$\left.\begin{array}{r} \sum Y_t = T\widehat{\alpha} + \widehat{\beta}\sum X_t \\ \sum Y_t X_t = \widehat{\alpha}\sum X_t + \widehat{\beta}\sum X_t^2 \end{array}\right\}$$

2 equation system with 2 unknowns!!

- Dividing the 1st. normal eq. by T:

$$\frac{1}{T}\sum Y_t = \frac{1}{T} T\widehat{\alpha} + \widehat{\beta}\frac{1}{T}\sum X_t$$

- That is:

$$\boxed{\widehat{\alpha}_{OLS} = \overline{Y} - \widehat{\beta}\,\overline{X}}$$

Estimation: Normal equations & LSE of β

- Substituting $\widehat{\alpha}$ in the 2nd. normal eq.:

$$\sum Y_t X_t = (\overline{Y} - \widehat{\beta}\overline{X})\sum X_t + \widehat{\beta}\sum X_t^2$$

- ...dividing by T and grouping together:

$$\frac{1}{T}\sum Y_t X_t = (\overline{Y} - \widehat{\beta}\overline{X})\frac{1}{T}\sum X_t + \widehat{\beta}\frac{1}{T}\sum X_t^2$$
$$\frac{1}{T}\sum Y_t X_t - \overline{Y}\,\overline{X} = \widehat{\beta}\left(\frac{1}{T}\sum X_t^2 - \overline{X}^2\right)$$

- ...and solving for the unknown:

$$\widehat{\beta} = \frac{\frac{1}{T}\sum Y_t X_t - \overline{Y}\,\overline{X}}{\frac{1}{T}\sum X_t^2 - \overline{X}^2} = \frac{\frac{1}{T}\sum y_t x_t}{\frac{1}{T}\sum x_t^2} \quad \begin{bmatrix}\text{Why?}\\ \text{Why?}\end{bmatrix} \quad \longrightarrow$$

- That is:

$$\boxed{\widehat{\beta}_{OLS} = \frac{\sum y_t x_t}{\sum x_t^2} = \frac{\text{Cov}(Y,X)}{\text{Var}(X)}}$$

Recall: variances and covariances?

- variance from original (uncentred) data?

$$\text{Var}(X) = \frac{1}{T}\sum x_t^2 = \frac{1}{T}\sum (X_t - \overline{X})^2$$
$$= \frac{1}{T}\sum X_t^2 + \frac{1}{T}\sum \overline{X}^2 - \frac{2}{T}\overline{X}\sum X_t$$

$$\boxed{\frac{1}{T}\sum x_t^2 = \frac{1}{T}\sum X_t^2 - \overline{X}^2}$$

- covariance from original (uncentred) data?

$$\text{Cov}(Y,X) = \frac{1}{T}\sum x_t y_t = \frac{1}{T}\sum (X_t - \overline{X})(Y_t - \overline{Y})$$
$$= \frac{1}{T}\sum X_t Y_t + \frac{1}{T}\sum \overline{X}\,\overline{Y} - \frac{1}{T}\overline{Y}\sum X_t - \frac{1}{T}\overline{X}\sum Y_t$$

$$\boxed{\frac{1}{T}\sum x_t y_t = \frac{1}{T}\sum X_t Y_t - \overline{X}\,\overline{Y}}$$

Numerical example: strawberry prod data

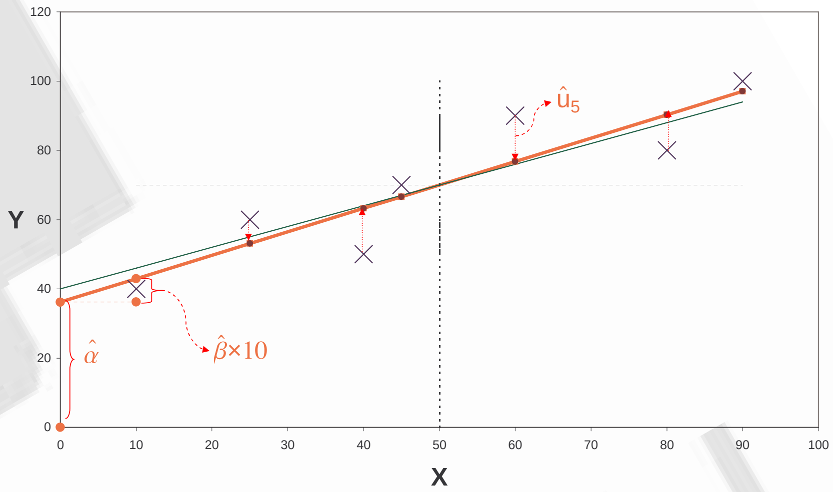
- Data...
- Centred data or "in deviation form"
  (deviations from respective means)...
- Squares and products...

	Y	X	y	x	y^2	x^2	yx
	40	10	-30	-40	900	1600	1200
	60	25	-10	-25	100	625	250
	50	40	-20	-10	400	100	200
	70	45	0	-5	0	25	0
	90	60	20	10	400	100	200
	80	80	10	30	100	900	300
	100	90	30	40	900	1600	1200
Sum					2800	4950	3350
Average	70	50	0	0	400	707.14	478.57

$\widehat{\alpha} = 36.162 \; (= \overline{Y} - \widehat{\beta}\overline{X})$ $\qquad \widehat{\beta} = 0.677 \left(= \frac{\text{Cov}(Y,X)}{\text{Var}(X)} \right)$

Can also use formulae based on original data... (Exercise: Try it!!)

Numerical example: strawberry regres plot



2.4a Properties of the Sample Regression Function.

Properties of residuals and SRF (1)

$\widehat{\beta}_{\text{OLS}} \rightsquigarrow \widehat{\alpha}_{\text{OLS}} \rightsquigarrow \widehat{Y}_t = \widehat{\alpha} + \widehat{\beta}X_t \rightsquigarrow \widehat{u}_t = Y_t - \widehat{Y}_t$

1. residuals add up to zero: $\sum \widehat{u}_t = 0$
   Demo: directly from 1st.o.c. □
2. $\overline{\widehat{Y}} = \overline{Y}$
   Demo: by def.: $\widehat{u}_t = Y_t - \widehat{Y}_t \rightsquigarrow \overline{\widehat{Y}} = \overline{Y} - \overline{\widehat{u}}$,
   but $\overline{\widehat{u}} = \frac{1}{T}\sum \widehat{u}_t = 0$ (from prop 1) $\rightsquigarrow \overline{\widehat{Y}} = \overline{Y}$. □
3. the SRF passes thru the pair of means $(\overline{X}, \overline{Y})$:
   $\overline{Y} = \widehat{\alpha} + \widehat{\beta}\overline{X}$
   Demo: from $\widehat{\alpha} = \overline{Y} - \widehat{\beta}\overline{X}$ (1st. normal eq.) □

Econometrics AFG

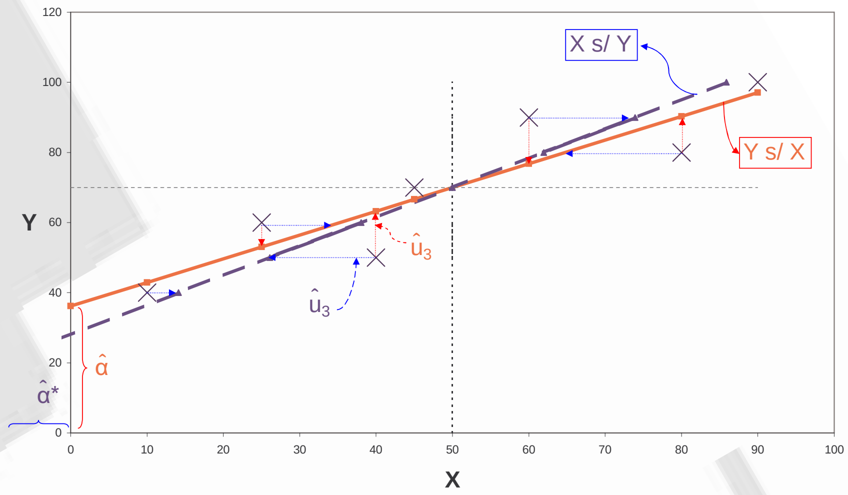
Properties of residuals and SRF (2)

4. residuals orthogonal to expl. v. X: $\sum X_t \widehat{u}_t = 0$
   Demo: directly from 1st.o.c.
5. residuals orthogonal to the explained part of Y: $\sum \widehat{Y}_t \widehat{u}_t = 0$
   Demo: $\sum(\widehat{\alpha} + \widehat{\beta} X_t)\, \widehat{u}_t =$

$$\widehat{\alpha} \underbrace{\sum \widehat{u}_t}_{=0 \text{ (from prop 1)}} + \widehat{\beta} \underbrace{\sum X_t \widehat{u}_t}_{=0 \text{ (from prop 4)}} = 0$$

Econometrics AFG

Causality: Y on X vs X on Y



Econometrics AFG

Properties of residuals and SRF (5)

8. $\widehat{\alpha}_{\text{OLS}}$ and $\widehat{\beta}_{\text{OLS}}$ unbiased $\rightsquigarrow$ expected value $=$ true value!
   Demo:

- $$\widehat{\beta} = \frac{\sum y_t x_t}{\sum x_t^2}$$
  $$\mathsf{E}(\widehat{\beta}) = \frac{1}{\sum x_t^2} \sum \underbrace{\mathsf{E}(y_t)}_{\beta x_t} x_t = \frac{1}{\sum x_t^2}\, \beta \sum x_t^2$$
  $$\boxed{\mathsf{E}(\widehat{\beta}) = \beta}$$
- $$\widehat{\alpha} = \overline{Y} - \widehat{\beta}\overline{X}$$
  $$\mathsf{E}(\widehat{\alpha}) = \frac{1}{T}\sum \mathsf{E}(Y_t) - \mathsf{E}(\widehat{\beta})\overline{X}$$
  $$= \frac{1}{T}\sum(\alpha + \beta X_t) - \beta\overline{X} = \alpha + \beta\overline{X} - \beta\overline{X}$$
  $$\boxed{\mathsf{E}(\widehat{\alpha}) = \alpha}$$

Econometrics AFG

2.5a Goodness of Fit: the Coefficient of Determination (R^2).

Econometrics AFG

Goodness of fit: Coefficient of determination

- Sum-of-Squares decomposition:

$$\sum Y_t^2 = \sum(\widehat{Y}_t^2 + \widehat{u}_t^2 + 2\widehat{Y}_t\widehat{u}_t)$$
$$= \sum \widehat{Y}_t^2 + \sum \widehat{u}_t^2 \quad \text{(from prop 5)}$$

- $\sum Y_t^2 - T\overline{Y}^2 = \sum \widehat{Y}_t^2 - T\overline{\widehat{Y}}^2 + \sum \widehat{u}_t^2$ (from prop 2)
- $$\boxed{\underset{(TSS)}{\sum y_t^2} = \underset{(ESS)}{\sum \widehat{y}_t^2} + \underset{(RSS)}{\sum \widehat{u}_t^2}}$$
- Definition of R^2:

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

$0 \leq R^2 \leq 1$ (Interpretation in terms of total variance??)

Econometrics AFG

No intercept $\rightsquigarrow$ invalid R^2

SLRM: $Y_t = \beta X_t + u_t$,

- apply Least-Squares fit:

$$\min_\beta \sum_{t=1}^T u_t^2 \quad \text{where} \quad u_t = Y_t - \beta X_t :$$

- First derivatives:

$$\frac{\partial \sum u_t^2}{\partial \beta} = 2\sum u_t \frac{\partial u_t}{\partial \beta} = 2\sum u_t(-X_t)$$

- 1st.o.c. (minimum) $\Rightarrow$ first derivative = zero:

$$\sum \widehat{u}_t X_t = \sum (Y_t X_t - \widehat{\beta} X_t^2) = 0$$

- $\nexists$ 1st equation!! $\rightsquigarrow$ $\begin{cases} \sum \widehat{u}_t \neq 0, \\ \overline{\widehat{Y}} \neq \overline{Y}, \end{cases}$ $\rightsquigarrow$ invalid R^2 (Why?)

Econometrics AFG

Relationship of R^2 with correlation coef

$$R^2 = \frac{\frac{1}{T}\sum \widehat{y}_t^2}{\frac{1}{T}\sum y_t^2} = \frac{\frac{1}{T}\sum(\widehat{\beta}x_t)^2}{\frac{1}{T}\sum y_t^2} = \frac{\widehat{\beta}^2 \frac{1}{T}\sum x_t^2}{\frac{1}{T}\sum y_t^2}$$
$$= \widehat{\beta}^2 \frac{\text{Var}(X)}{\text{Var}(Y)} = \frac{\text{Cov}(Y,X)^2}{\text{Var}(X)^2}\frac{\text{Var}(X)}{\text{Var}(Y)}$$
$$= \frac{\text{Cov}(Y,X)^2}{\text{Var}(X)\text{Var}(Y)}$$

$R^2 = r^2_{X,Y}$

Econometrics AFG

Numerical example: strawberry prod data (cont)

- recall data & previous calculations. . .
- do the same for fitted values. . .
- now calculate R^2. . .

	y^2	$\widehat{Y}$	$\widehat{y}$	$\widehat{y}^2$	$\widehat{u}$	$\widehat{u}^2$
	900	42.92	-27.07	732.82	-2.92	8.58
	100	53.08	-16.91	286.25	6.91	47.87
	400	63.23	-6.76	45.80	-13.23	175.09
	0	66.61	-3.38	11.45	3.38	11.45
	400	76.76	6.76	45.80	13.23	175.09
	100	90.30	20.30	412.21	-10.30	106.15
	900	97.07	27.07	732.82	2.92	8.58
Average	400	70	0	323.88		
Sum	2800			2267.17		532.82
	TSS			ESS		RSS

$R^2 = 0.8097$ ($= \frac{\text{ESS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}}$)

(Exercise: How does this compare with Corr(X,Y)? . . . Try it!!)

Econometrics AFG

2.3b OLS in the GLRM.

GLRM: the PRF

- Recall: model with K explanatory variables:

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

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.
  - $\beta_k = \dfrac{\partial E(Y_t)}{\partial X_{kt}} \simeq \dfrac{\Delta E(Y_t)}{\Delta X_{kt}}, \quad k = 1 \ldots K$: 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 $\widehat{\beta} = (\widehat{\beta}_0, \widehat{\beta}_1 \ldots, \widehat{\beta}_K)'$ 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:

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

- or in matrix form:
  [recall:

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

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

- that is

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

Note: vector derivatives

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

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

- With vectors and matrices this is quite similar:

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

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

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 $u'u$

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

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 (co

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.: $X'\widehat{u} = 0$
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):

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

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

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

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

$$0 \leq R^2 \leq 1$$

Goodness of fit: R^2 Revisited (cont)

Note 1: R^2 measures the proportion of the dependent variable variation explained by (a linear combination) of the variations 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 $\mathrm{Var}(u_t)$

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

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:

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

which is unbiased:

$$\text{E}(\widehat{\sigma}^2) = \sigma^2.$$

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

Properties of the Least-Squares 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 $\text{E}(u) = 0$, $\widehat{\beta}_{\text{OLS}}$ is unbiased:

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

Properties of the Least-Squares Estimator (2)

- **Variance:** Recall:

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

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

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

Properties of the Least-Squares Estimator (2con

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

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

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

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

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:
(SEE NOTES)

2.3c OLS: Main 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

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

2.7a Omission of relevant variables.

Omission of relevant variables

- true relationship:

$$Y = X\beta + u = \begin{bmatrix} X_I & \vline & X_{II} \end{bmatrix} \begin{pmatrix} \beta_I \\ \hline \beta_{II} \end{pmatrix} + u$$

$$X = \left[\begin{array}{cccc|ccc} 1 & x_{11} & \dots & x_{K_1,1} & x_{K_1+1,1} & \dots & x_{K1} \\ 1 & x_{12} & \dots & x_{K_1,2} & x_{K_1+1,2} & \dots & x_{K2} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & x_{1T} & \dots & x_{K_1,T} & x_{K_1+1,T} & \dots & x_{KT} \end{array}\right], \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{K_1} \\ \hline \beta_{K_1+1} \\ \vdots \\ \beta_K \end{pmatrix}.$$

$$Y = X_I\beta_I + X_{II}\beta_{II} + u$$

- estimated relationship:

$$Y = X_I\beta_I + v \qquad \text{where} \quad v = X_{II}\beta_{II} + u,$$

then $\mathsf{E}(v) \neq 0 \quad \rightsquigarrow \quad \mathsf{E}(\widehat{\beta}) \neq \beta$.

i.e. $\widehat{\beta}$ is biased.

Omission of relevant variables: consequences

Summary:

- OLS estimator of coefficients is *biased* (except if $x'_I x_{II} = 0$).
- OLS estimator of intercept is *always biased*.
- Estimator of Error variance is *always biased*.

2.7b Multicollinearity

Perfect Multicollinearity

Extreme case:

- exact linear combination:
  - $\sum_{k=0}^{K} \lambda_k X_{kt} = 0, \quad \lambda \neq 0, \quad X_{0t} = 1,$
  - $\exists X_i \mid X_i = \lambda_0^* + \sum_{k=1, k \neq i}^{K} \lambda_k^* X_{kt},$
  - $\exists X_i, X_j \mid$ Corr$(X_i, X_j) = 1,$
  - $\exists X_i \mid$ aux regres X_i on $\{X_k\}_{k=1, k\neq i}^{K} \quad \rightsquigarrow \quad \mathrm{R}_i^2 = 1.$
- Problem:
  - $\mathrm{rk}\, X < K+1$, (X isn't of full rank)
  - $\rightsquigarrow \quad \det(X) = 0$
  - $\rightsquigarrow \quad \nexists (X'X)^{-1}$
  - $\rightsquigarrow \quad \widehat{\beta}$?

Perfect Multicollinearity: example

- Let $X_{4t} = 2X_{1t} \quad \forall t$:

$$X_{4t} = 0 + 2X_{1t} + 0 \cdot X_{2t} + 0 \cdot X_{3t} + 0 \cdot X_{5t} + \cdots + 0 \cdot X_{Kt},$$

- no error? $\Rightarrow$ aux regres X_4 on $\{X_k\}_{k=1, k\neq 4}^{K} \quad \rightsquigarrow \quad \mathrm{R}_4^2 = 1$!!
- Model specification:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \cdots + u_t, t = 1, 2 \ldots, T,$$
$$X_{4t} = 2X_{1t},$$

- and substituting in model:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 (2X_{1t}) + \cdots + u_t,$$
$$= \beta_0 + \underbrace{(\beta_1 + 2\beta_4)}_{\beta_1^\star} X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \cdots + u_t$$

- now we have one less parameter to estimate.

Multicollinearity: counterexample

$$Y_t = \beta_0 + \beta_1^\star X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \cdots + u_t$$

- Just K parameters remain to be estimated,
  but β_1 and β_4 cannot be estimated separately:
  - we can just estimate a linear combination of them:
    $\beta_1^\star = \beta_1 + 2\beta_4,$
  - *i.e.* combined effect of X_{1t} and X_{4t} on Y_t!!
- (Exercise: Try it yourself with $X_{2t} - 3X_{3t} = 10, \quad \forall t$.)
- multicollinearity = *linear relationships*
  but... what if relationship *isn't linear*? *e.g.:*

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{1t}^2 + u_t$$

  - X is of full column rank $\rightsquigarrow$ no problem.

Perfect Multicollinearity: consequences

- some parameters cannot be estimated separately.
- some estimates are just l.c. of parameters.
- R^2 is correct:
  correctly picks up proportion of Y_t explained by the regression.
- Predictions of Y are still valid.

2.7c Imperfect Multicollinearity

Imperfect Multicollinearity

- Problem:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \cdots + u_t, t = 1, 2 \ldots, T,$$

$$X_{4t} = 2X_{1t} + v_t,$$

$v_t =$ gap between X_{4t} and $2\,X_{1t}$,

- approximate relationship:
- auxiliary regression X_{4t} on rest $\rightsquigarrow$ $R^2 \approx 1$.
- it's a matter of degree ($x'x$ not diagonal $\rightsquigarrow$ correlated variables)
- Note: whenever perfect/imperfect is not specified we mean imperfect mc.

Multicollinearity: Symptoms

- Typical symptom:
  - high R^2 (relevant group of regressors)
  - but "t" ratios not significant (inability to separate effects of regressors).
- more formally:

$$\text{Var}(\widehat{\beta}^\star) = \sigma^2 (x'x)^{-1} = \frac{\sigma^2}{T} \text{Var}(X^\star)^{-1}$$

$$\Rightarrow \quad \text{Var}(\widehat{\beta}_k) = \frac{\sigma^2}{T \text{Var}(X_k)(1 - R_k^2)},$$

- so that, in the previous example $X_{4t} \approx 2X_{1t}$:
  - Corr$(X_4, X_1) \uparrow$
  - R_4^2 and R_1^2 $\upuparrows$
  - denominator $\downarrow$
  - variances $\upuparrows$

Multicollinearity: Consequences

- Some coefficients aren't significant, even if their variables have an important effect on dependent variable.
- Nevertheless, Gauss-Markov $\Rightarrow$ linear, unbiased and of minimum variance estimators, then *it isn't possible to find a Better LUE.*
- R^2 is correct: correctly picks up proportion of Y_t explained by the regression.
- Predictions of Y are still valid.

Multicollinearity: How to detect

- Small changes in data
  ⇒ important changes in estimates
  (they can even affect their signs).
- Coefficient estimations
  not individually significant.
- Regressors are jointly significant.
- High coefficient of determination R^2.
- Auxiliary regressions among regressors
  ⇒ high R_k^2.

Multicollinearity: Some solutions

Multicollinearity is not an easy problem to solve.
Nevertheless, from

$$\text{Var}(\widehat{\beta}_k) = \frac{\sigma^2}{T\text{Var}(X_k)(1 - R_k^2)},$$

it turns that:

- $\mathbf{T}\uparrow$: Increase number of observations T.
  Also, differences among regressors may increase.
- $\text{Var}(\mathbf{X})\uparrow$: *e.g.* study about consumption function:
  sample of families ⇝ all possible rents.
- $\text{Var}(\mathbf{X})\uparrow$: Additional information.
  e.g. impose restrictions suggested by Ec. Th.
- $\sigma^{\mathbf{2}}\downarrow$: New relevant regressor not yet included.
  It would also avoid serious bias problems.
- $\mathbf{R_k^2}\downarrow$: Eliminate variables that may produce multicollinearity.
  (Take care of omitting some relevant regressor though).

2.8 The Least-Squares Estimator under Restrictions.

GLRM under linear restrictions (1)

- previous chapter objectives:
  - Econometric model (GLRM), characteristics and basic assumptions. . .
  - but. . . no knowledge about model parameters.
  - Least Squares Method for parameter estimation (OLS).
  - Properties of resulting estimators.
- present chapter objectives:
  - *a priori* information about parameter values (or l.c.) . . .
  - given by
    - economic theory,
    - other empirical work,
    - own experience, etc.
  - Non-Restricted Model ⇒ Ordinary LS.
  - Restricted Model ⇒ Restricted LS.
  - Check, given the estimated model, if they are compatible with available data.

GLRM under linear restrictions: examples

- production function with constant returns to scale: $\beta_K + \beta_L = 1$.
- product demands as function of price: $\beta = -1$ (say).
- in GLRM: let us assume that $\beta_2 = 0$ and $2\beta_3 = \beta_4 - 1$:
  - Full model:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_{Kt} X_{Kt} + u_t, \text{ with } \beta_2 = 0 \text{ and } 2\beta_3 + 1 = \beta_4;$$

  - Alternative transformed model:

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 X_{1t} + 0 X_{2t} + \beta_3 X_{3t} + (2\beta_3 + 1) X_{4t} + \cdots + \beta_{Kt} X_{Kt} + u_t \\ Y_t - X_{4t} &= \beta_0 + \beta_1 X_{1t} + \beta_3 (X_{3t} + 2X_{4t}) + \cdots + \beta_K X_{Kt} + u_t \\ Y_t^* &= \beta_0 + \beta_1 X_{1t} + \beta_3 Z_t + \cdots + \beta_K X_{Kt} + u_t \end{aligned}$$

    where $Y_t^* = Y_t - X_{4t}$ and $Z_t = X_{3t} + 2X_{4t}$.
  - This transformed model:
    - can be estimated by OLS:
      $\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_3, \widehat{\beta}_5, \ldots, \widehat{\beta}_K$, together with $\widehat{\beta}_2 = 0$ and $\widehat{\beta}_4 = 2\widehat{\beta}_3 + 1$.
    - has new endogenous variable Y_t^* (not always so: e.g. if $\beta_2 = 0$ alone) and new explanatory variable Z_t.

GLRM under linear restrictions (2)

- The "transformation" method is good for simple cases only.
- In general, q (nonredundant) linear restrictions among parameters:

$$\begin{matrix} 1 \\ \vdots \\ q \end{matrix} \begin{pmatrix} \diamond & \diamond & \diamond & \ldots & \diamond \\ \vdots & & & & \\ \diamond & \diamond & \diamond & \ldots & \diamond \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \cdots \\ \beta_K \end{pmatrix} = \begin{pmatrix} \diamond \\ \vdots \\ \diamond \end{pmatrix}$$

  - for given matrix R and vector r,

$$\underset{(q \times K+1)}{R} \; \beta = \underset{(q \times 1)}{r}$$

  - example of non-valid case (why?):

$$\beta_3 = 0, \quad 2\beta_2 + 3\beta_4 = 1, \quad \beta_1 - 2\beta_4 = 3, \quad 6\beta_4 = 2 - 4\beta_2 + \beta_3$$

GLRM under linear restrictions (2cont)

- Write previous example $\beta_2 = 0$ and $2\beta_3 = \beta_4 - 1$ ($q = 2$ restrictions) as in general formula:

$$\underset{(2 \times K+1)}{\underset{R}{\begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & 2 & -1 & 0 & \ldots & 0 \end{pmatrix}}} \underset{(K+1 \times 1)}{\underset{\beta}{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \cdots \\ \beta_K \end{pmatrix}}} = \underset{(2 \times 1)}{\underset{r}{\begin{pmatrix} 0 \\ -1 \end{pmatrix}}} .$$

- In general, we write GLRM subject to q linear restrictions as:

$$\underset{(T \times 1)}{Y} = \underset{(T \times K+1)}{X} \underset{(K+1 \times 1)}{\beta} + \underset{(T \times 1)}{u} ,$$

$$\underset{(q \times K+1)}{R} \underset{(K+1 \times 1)}{\beta} = \underset{(q \times 1)}{r} .$$

Estimation: restricted least squares (RLS).

- Typical optimization exercise:

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

  subject to $R\beta = r$.

- Lagrangian:

$$L(\beta, \lambda) = u'u - 2\lambda'(R\beta - r)$$

$$\min_{\beta, \lambda} L(\beta, \lambda).$$

- First derivatives:

$$\frac{\partial L(\beta, \lambda)}{\partial \beta} = -2\, X'u - 2\, R'\lambda,$$

$$\frac{\partial L(\beta, \lambda)}{\partial \lambda} = -2\,(R\beta - r),$$

Estimation: restricted least squares (RLS) (cont)

- 1st.o.c. $\rightsquigarrow$ normal equations:

$$X' \hat{u}_R + R' \hat{\lambda} = 0, \quad (4)$$

$$R \hat{\beta}_R = r, \quad (5)$$

where $\hat{\beta}_R$ and $\hat{\lambda}$ are values of β, λ that satisfy 1st.o.c. and residuals

$$\hat{u}_R = Y - X \hat{\beta}_R. \quad (6)$$

- Solving for $\hat{\beta}_R$: (see notes)

$$\hat{\beta}_R = \hat{\beta} + A(r - R\hat{\beta}) = (I - AR)\,\hat{\beta} + Ar \quad (7)$$

where $A = (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}$.

RLS estimation: characteristics

- Expression (7): $\hat{\beta}_R = \hat{\beta} + A(r - R\hat{\beta}) \rightsquigarrow$
  - the restricted estimate $\hat{\beta}_R$ can be obtained as a function of the (not restricted) ordinary estimate: $\hat{\beta}$
  - $R\hat{\beta} \simeq r \quad \Rightarrow \quad \hat{\beta}_R$ (restricted) $\simeq \hat{\beta}$ (not restricted) .
- Normal equations (4): $X' \hat{u}_R + R' \hat{\lambda} = 0 \rightsquigarrow$
  - satisfy the restrictions (obvious).
  - $X'\hat{u}_R \neq 0$, *i.e.*:
    - sum of restricted residuals not zero,
    - restricted residuals not orthogonal to explanatory variables,
    - then, restricted residuals not orthogonal to fitted $\hat{Y}_R$.
  - TSS $\neq$ RSS$_R$ + ESS$_R$
    (compare with ordinary case and with transformed equation: R^2 ??).

Properties of the RLS estimator (1)

Expression (7) : $\hat{\beta}_R = (I - AR)\,\hat{\beta} + Ar \quad \rightsquigarrow$

1. **Linear:** RLS estimator $\hat{\beta}_R$ is l.c. of OLS estimator $\hat{\beta}$, which is linear , then $\hat{\beta}_R$ is linear also .
2. **Bias:** RLS estimator $\hat{\beta}_R$ is $\begin{cases} \text{biased,} & \text{if } R\,\beta \neq r\ , \\ \text{unbiased,} & \text{if } R\,\beta = r \text{ true} \end{cases}$

   Demo:

$$\text{E}(\hat{\beta}_R) = (I - AR)\,\text{E}(\hat{\beta}) + Ar = (I - AR)\,\beta + Ar = \beta + A(r - R\beta).$$

3. **Covariance Matrix:** $\text{Var}(\hat{\beta}_R) = (I - AR)\text{Var}(\hat{\beta}) = \sigma^2(I - AR)(X'X)^{-1}$
   Demo: (see notes)

Properties of the RLS estimator (2)

4. **Smaller variance** than OLS estimators,
   even if restrictions aren't true:
   Demo:

$$\text{Var}(\hat{\beta}_R) = \text{Var}(\hat{\beta}) - AR\,\text{Var}(\hat{\beta}) = \text{Var}(\hat{\beta}) - (\text{psd matrix}).$$

□

5. surprising result (apparently):
   - less "uncertainty" about parameters
     $\rightsquigarrow$ greater precision in estimation. . .
   - . . . but towards an erroneous result (biased)
     if restriction isn't true.

Multicollinearity vs restrictions

Must clearly distinguish two different cases:

- linear relationships among regressors (*i.e.* multicollinearity):

  e.g. $X_{4t} = 2X_{1t}$

  $\Rightarrow$ missing information for individual estimates.

- linear relationships among coefficients:

  e.g. $\beta_4 = 2\,\beta_1$

  $\Rightarrow$ extra information about parameters

  $\rightsquigarrow$ estimators with smaller variance.

- respective models to estimate:

$$Y_t = \beta_0 + (\underbrace{\beta_1 + 2\beta_4}_{\beta_1^*})X_{1t} + \beta_2 X_{2t} + + \cdots + u_t,$$

$\Rightarrow \widehat{\beta}_1^*$ but $\widehat{\beta}_1, \widehat{\beta}_4$?

$$Y_t = \beta_0 + \beta_1(\underbrace{X_{1t} + 2X_{4t}}_{X_{1t}^*}) + \beta_2 X_{2t} + + \cdots + u_t,$$

$\Rightarrow \widehat{\beta}_1$ and $\widehat{\beta}_4 = 2\,\widehat{\beta}_1$