ECONOMETRICS II

Topic 1: Brief Review of the General Linear Model and Asymptotic Results

Luca Salvadori

UAB, BSE, TARC, IEB

Academic Year 22-23

1. Brief Review of the General Linear Model (GLM)
Econometrics I

NOTATION:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_k X_{kt} + u_t, \qquad t = 1, 2, \ldots, T$$

VECTORIAL NOTATION

$$Y_t = X_t^{'} \beta + u_t$$

$$t = 1, 2, \ldots, T$$

$$X_t = (1, X_{2t}, \ldots, X_{kt})^{'}$$

1. Brief Review of the General Linear Model (GLM)
Econometrics I

MATRIX FORM NOTATION:

$$Y = X\beta + u$$

$$\underset{(T\times 1)}{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_T \end{bmatrix}, \underset{(k\times 1)}{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{bmatrix}, \underset{(T\times 1)}{u} = \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_T \end{bmatrix}$$

$$\underset{(T\times k)}{X} = \begin{bmatrix} 1 & X_{21} & \dots & X_{k1} \\ 1 & X_{22} & \dots & X_{k2} \\ \dots & \dots & \dots & \dots \\ 1 & X_{2T} & \dots & X_{kT} \end{bmatrix}$$

1. Brief Review of the General Linear Model (GLM)
Econometrics I

BASIC HYPOTHESES:

1. $X_1, \ldots X_k$ and y are **iid.**

2. $r(X) = k$; x_i are linearly independent
   - X has full rank, thus $\exists\ (X'X)^{-1}$
   - **there is not perfect multicollinearity**;

3. X_t and u_t have nonzero finite fourth order moments.

4. $E(u_t|\mathrm{X})=0 \Rightarrow E(u_t) = 0,\ t = 1,2,\ \ldots,\ k$

1. Brief Review of the General Linear Model (GLM)
Econometrics I

BASIC HYPOTHESES:

5. $Var(u_t) = E(u^2) = \sigma^2$, $t = 1,2, \ldots, T$ (**homoscedasticity**)
   $Cov(u_t, u_s) = E(u_t u_s) = 0$, $\forall\ t \neq s$ (***u_t* not serially correlated**)

$$Var(u) = E(uu') = \sigma^2 I_T .$$

6. **Normality:** $u \sim N(0, \sigma^2 I)$; $u_t \sim N(0, \sigma^2)$

1. Brief Review of the General Linear Model (GLM)
Econometrics I

INTERPRETATION OF THE COEFFICIENTS:

Given the population expectation of Y_t, $E(Y_t) = b_1 + b_2 X_{2t} + \ldots + b_k X_{kt}$:

$$\beta_j = \frac{\partial E(Y_t)}{\partial X_{jt}}, \qquad \text{for any} \qquad j = 2, \ldots, k$$

It measures the expected change in Y if X_j changes by one unit (*ceteris paribus*, that is, keeping the value of the other explanatory variables constant). It's the marginal effect of X_j on Y.

1. Brief Review of the General Linear Model (GLM)
Econometrics I

BASIC RESULTS FOR THE OLS ESTIMATION:

- $\widehat{\beta} = (X'X)^{-1}X'Y = [(X'X)^{-1}X'][X\beta + u] = \beta + (X'X)^{-1}X'u$
- $E(\widehat{\beta}) = \beta + E[(X'X)^{-1}X'u] \underset{\text{given that } E(u)=0}{=} \beta$

  Therefore $\widehat{\beta}$ are unbiased without the need for assumptions 5-6 to hold, only 1-4 are needed

- $Var[\widehat{\beta}] = \sigma^2(X'X)^{-1}; \widehat{Var}[\widehat{\beta}] = \widehat{\sigma}^2(X'X)^{-1}$ $\qquad \hat{\sigma}_u^2 = \dfrac{e'e}{N-k}$ $, e = \hat{u}$
- Standard errors: $SE(\widehat{\beta}_j) = \sqrt{\widehat{\sigma}^2(X'X)^{-1}_{jj}}, j = 1, \ldots, k$

1. Brief Review of the General Linear Model (GLM)
Econometrics I

BASIC RESULTS FOR THE OLS ESTIMATION:

- Estimated/adjusted values/predictions

$$\widehat{Y}_t = \widehat{\beta}_1 + \widehat{\beta}_2 X_{2t} + \ldots + \widehat{\beta}_k X_{kt}, t = 1, 2, \ldots, T \quad (2)$$

- Residuals $e_t = Y_t - \widehat{\beta}_1 - \widehat{\beta}_2 X_{2t} - \ldots - \widehat{\beta}_k X_{kt},$

$t = 1, 2, \ldots, T.$ In vectorial notation: $\underset{(T\times 1)}{e} = \begin{bmatrix} e_1 \\ e_2 \\ \ldots \\ e_T \end{bmatrix}$

1. Brief Review of the General Linear Model (GLM)
Econometrics I

BASIC RESULTS FOR THE OLS ESTIMATION:

- Interpretation of estimated values

$$\widehat{\beta}_j = \frac{\partial \widehat{Y}}{\partial X_j}, \quad \text{to} \quad j = 2, \ldots, k$$

Therefore it measures the *marginal* effect that an increase by one unit in X_J will have on the dependent variable, keeping the remaining explanatory variables of the model constant. In the linear regression model **marginal effects are constant.**

1. Brief Review of the General Linear Model (GLM)
Econometrics I

BASIC RESULTS FOR THE OLS ESTIMATION:

- Some common alternatives to the estimated model (2)

1. <u>Model in logarithms (log-log)</u>

$$\widehat{\log(Y_t)} = \widehat{\beta}_1 + \widehat{\beta}_2 \log(X_{2t}) + \ldots + \widehat{\beta}_k \log(X_{kt})$$

The $\widehat{\beta}_j,\ j = 2, \ldots, k$ are the estimated elasticities, that is, $\widehat{\beta}_j$ measures the estimated variation in percentage points for the dependent variable if the explanatory variable increases by 1%. **Elasticities are constant.**

1. Brief Review of the General Linear Model (GLM)
Econometrics I

BASIC RESULTS FOR THE OLS ESTIMATION:

- Some common alternatives to the estimated model (2)

2. <u>Model in semi-logarithms (log-level)</u>

$$\widehat{\log(Y_t)} = \widehat{\beta}_1 + \widehat{\beta}_2 X_{2t} + \ldots + \widehat{\beta}_k X_{kt}$$

For any $j = 2, \ldots, k,\ 100^* \widehat{\beta}_j$ measures the estimated variation in p.p. for the dependent variable if the explanatory variable increases by 1 unit.**The semi-elasticitiesare constant**.

1. Brief Review of the General Linear Model (GLM)
Econometrics I

BASIC RESULTS FOR THE OLS ESTIMATION:

- Some common alternatives to the estimated model (2)

3. <u>Model in semi-logarithms (level-log)</u>

$$Y_t = \widehat{\beta}_1 + \widehat{\beta}_2 \log(X_{2t}) + \ldots + \widehat{\beta}_k \log(X_{kt})$$

For any $j = 2, \ldots, k$, $\widehat{\beta}_j$/100 measures the estimated variation in units for the dependent variable if the explanatory variable increases by 1%.

1. Brief Review of the General Linear Model (GLM)
Econometrics I

COEFFICIENT OF DETERMINATION (goodness-of-fit): $\boxed{R^2 = 1 - \frac{RSS}{TSS} = \frac{ESS}{TSS}}$

Given the model: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + u_i$

$$RSS \equiv \sum_{i=1}^{n} \hat{u}_i^2 \qquad TSS \equiv \sum_{i=1}^{n} (y_i - \bar{y})^2 \qquad ESS \equiv \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

ADJUSTED COEFFICIENT OF DETERMINATION :

$$\boxed{\bar{R}^2 = 1 - \frac{n-1}{n-k-1}(1 - R^2)}$$

1. Brief Review of the General Linear Model (GLM)
Econometrics I

INFERENCE

Given the model $$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + u_i$$

- We want to verify the reliability of statements about the parameters
- ... Particularly we want to find out if the variables x_j are relevant factors for the variable y

How?

1) t-test

2) Confidence intervals

3) F-test

1. Brief Review of the General Linear Model (GLM)
Econometrics I

INFERENCE: t-test

$$\hat{t} = \frac{\hat{\beta}_j - \bar{\beta}_j}{\hat{se}(\hat{\beta}_j)} \sim t_{n-k-1} \qquad \hat{se}(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 (X'X)^{-1}_{j+1,j+1}}$$

Elements of the test:

- ✓ Null hypothesis H_0: hypothesis that is assumed to be true at the beginning of the test
- ✓ Alternative hypothesis H_1: hypothesis that is valid if the null is not valid.
- ✓ example: *contrast of individual significance*
  - ❑ H_0: $\beta_1 = 0$
  - ❑ H_1: $\beta_1 \neq 0$ (two-tailed test)
  - ❑ H_1: $\beta_1 < 0$ or H_1: $\beta_1 > 0$ (one tail test)

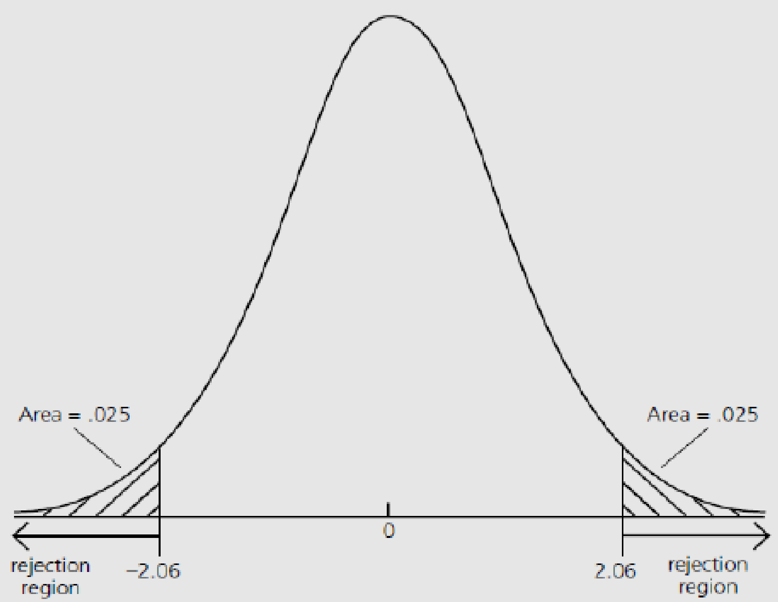
1. Brief Review of the General Linear Model (GLM)
Econometrics I

INFERENCE: t-student contrast

- $H_0 : \beta_j = \bar{\beta}_j$
- $H_1 : \beta_j \neq \bar{\beta}_j$
- Select α
- $$\hat{t} = \frac{\hat{\beta}_j - \bar{\beta}_j}{\widehat{se}(\hat{\beta}_j)}$$
- Reject if $|\hat{t}| > t_{n-k-1,\alpha/2}$

Figure 4.4

5% rejection rule for the alternative H_1: $\beta_j \neq 0$ with 25 *df*.



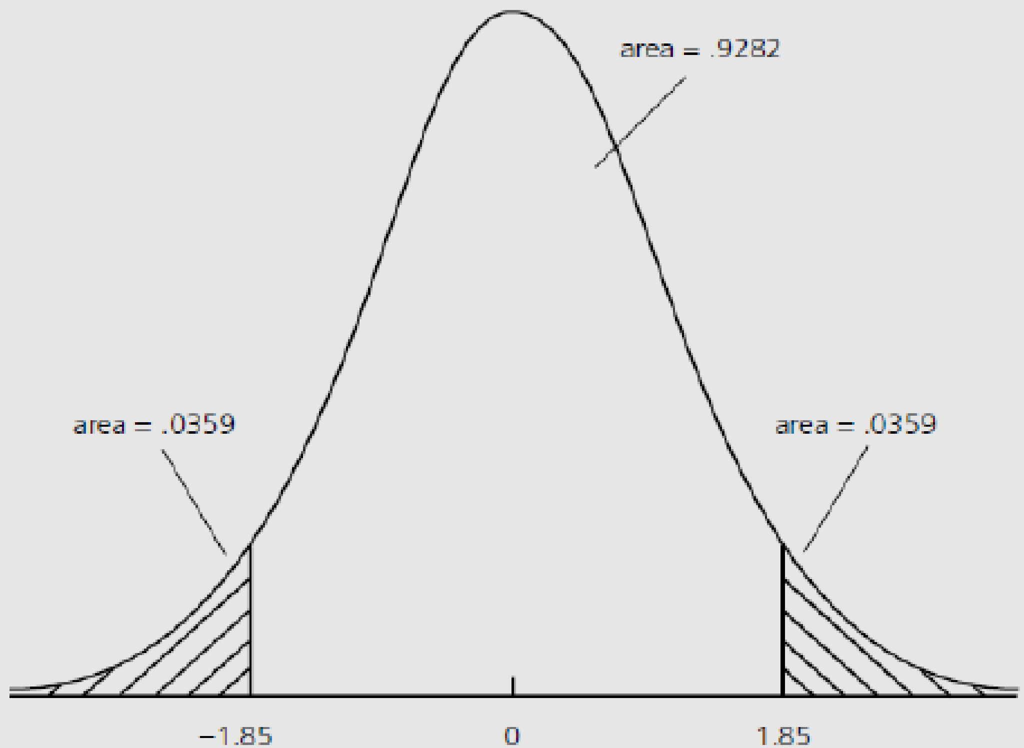
ECONOMETRICS II

1. Brief Review of the General Linear Model (GLM)
Econometrics I

INFERENCE: p-value

Figure 4.6

Obtaining the *p*-value against a two-sided alternative, when $t = 1.85$ and $df = 40$.



- Two-tailed test:

$$valor - p = p(|t_{n-k-1}| > |\hat{t}|)$$

- One tailed test:

$$valor - p = p(t_{n-k-1} > |\hat{t}|)$$

1. Brief Review of the General Linear Model (GLM)
Econometrics I

INFERENCE: confidence interval

Confidence interval $1-\alpha$

$$\hat{\beta}_j \pm \hat{se}(\hat{\beta}_j)\ t_{n-k-1,\alpha/2}$$

where

$$\hat{se}(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 (X'X)^{-1}_{j+1,j+1}}$$

1. Brief Review of the General Linear Model (GLM)
Econometrics I

INFERENCE: F - Test

We need to compare two models, the restricted model (R) and the unrestricted model (NR).

$$H_0 : \beta_2 = 0, \beta_3 = 0$$

NR $$M1 : \quad y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + u_i$$

R $$M2 : \quad y_i = \beta_0 + \beta_1 x_{i1} + \beta_4 x_{i4} + u_i$$

F-statistic: $$\hat{F} = \frac{(SCR_r - SCR)/q}{(SCR)/(n-k-1)}$$

q= number of restrictions

We reject H_o if $F > F_{q,n-k-1,\alpha}$ where $F_{q,n-k-1,\alpha}$ is the critical value of the F distribution with q and n-k-1 degrees of freedom and significance level α

2. Asymptotic theory
Statistics II

Asymptotic properties:

The distribution of an estimator can change with the size of the sample,we will analyze the properties of the estimators of the parameters in the MLG for large samples.

The asymptotic theory analyzes the approximate behavior of an estimator or statistic when $T \rightarrow \infty$

2. Asymptotic theory
Statistics II

• **Convergence in probability:** let $Z_1Z_2;....Z_T$ be a succession of random variables. We will say that it converges in probability to a constant "c", if:

$$\forall \varepsilon > 0 \lim_{T\to\infty} \Pr[|Z_T - c| > \varepsilon] = 0 \quad \text{or} \quad Z_T \to {}_p c \quad \text{or} \quad plim Z_T = c$$

• **Convergence in distribution**: let Z_t be a sequence of random vectors with distribution function F_t. We will say that Z_t converges in distribution to the vector Z if:

$$\lim_{T\to\infty} F_T(z) = F(z) \; \forall z$$
$$Z_T \to {}_d Z$$

where F is the distribution function of Z

2. Asymptotic theory statisticsII

•**Consistency:** an estimator $\widehat{\theta}$, function of a sample of size T, is a consistent estimator of the unknown parameter θ if by increasing the sample size T, the sequence $\{\hat{\theta}_T\} = \{\hat{\theta}_1, \hat{\theta}2, \ldots\}$ converges in probability to the true value of the parameter θ.

$$\forall \epsilon > 0 \qquad \lim_{T \to \infty} Pr\{|\hat{\theta}_T - \theta| < \epsilon\} = 1$$

$$\hat{\theta}_T \xrightarrow{p} \theta \qquad \text{ó como} \qquad \text{plim}_T \, \hat{\theta} = \theta$$

❖ Interpretation: if an estimator is consistent, as the size of the sample increases we are reducing the probability that the estimator will move away from the true value of the parameter that we want to estimate.

2. Asymptotic theory
Statistics II

• **Asymptotic unbiasedness:** $\hat{\theta}$ is an asymptotically unbiased estimator of the unknown parameter θ, if

$$\lim_{T \to \infty} E(\hat{\theta}_T) = \theta$$

- An unbiased estimator is also asymptotically unbiased
- But an asymptotically unbiased estimator may be biased

consistency $\not\Leftrightarrow$ **asymptotic unbiasedness**

• **Markov theorem**: If $\hat{\theta}_t$ it is an asymptotically unbiased estimator of θ and its variance converges to zero when T $\to\infty$, then $\hat{\theta}_t$ is a consistent estimator of θ.

2. Asymptotic theory
Statistics II

• **Asymptotic normality:** Consider $\hat{\theta}$ estimator of θ. $\hat{\theta}_t$ is asymptotically normal with bounded variance Ω (being Ω onepositive definite matrix if):

$$\sqrt{T}(\widehat{\theta}_T - \theta) \rightarrow {}_d N(0, \Omega)$$ or

$$\sqrt{T}(\widehat{\theta}_T - \theta) \simeq N(0, \Omega)$$

Using the properties of the Normal distribution

$$(\widehat{\theta}_T - \theta) \simeq N(0, \Omega/T) \qquad \widehat{\theta}_T \simeq N(\theta, \Omega/T)$$

If Ω is unknown, a consistent estimator can be obtained $\widehat{\Omega}$ and we can approximate the distribution of the estimator in this way:

$$\widehat{\theta}_T \simeq N(\theta, \widehat{\Omega}/T)$$

2. Asymptotic theory
Statistics II

- **Asymptotic efficiency:** Consider $\hat{\theta}$ estimator of θ. $\hat{\theta}_t$ is asymptotically efficient if:
    - It is asymptotically normal
    - and its variance is less than or equal to the limit of the variance of any other asymptotically normal estimator of the vector of parameters θ.

2. Asymptotic theory

ASYMPTOTIC PROPERTIES OF THE OLS ESTIMATOR

- **Asymptotic Unbiasedness**: Given that the OLS estimators of β and σ^2 are unbiased (via basic assumptions), they are also asymptotically unbiased

- **Consistency of $\hat{\boldsymbol{\beta}}$**: if the assumptions 1-4 of MLG are met then the OLS estimator of β is consistent (via LLN)

$$plim\hat{\beta} = plim[(X'X)^{-1}X'Y] = \beta + plim[(X'X)^{-1}X'u]$$
$$= \beta + plim\left[\left(\frac{X'X}{T}\right)^{-1}\frac{X'u}{T}\right] = \beta$$

- ➢ $plim\frac{X'X}{T} = E(X'X) = Q$
- ➢ $plim\frac{X'X}{T} = Q^{-1}$
- ➢ $plim\frac{X'u}{T} = E(X'u) = 0$

2. Asymptotic theory

The Multivariate Central Limit Theorem

KEY CONCEPT 18.2

Suppose that $\boldsymbol{W}_1, \ldots, \boldsymbol{W}_n$ are i.i.d. m-dimensional random variables with mean vector $E(\boldsymbol{W}_i) = \mu_W$ and covariance matrix $E[(\boldsymbol{W}_i - \mu_W)(\boldsymbol{W}_i - \mu_W)'] = \boldsymbol{\Sigma}_W$, where $\boldsymbol{\Sigma}_W$ is positive definite and finite. Let $\overline{\boldsymbol{W}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{W}_i$. Then $\sqrt{n}(\overline{\boldsymbol{W}} - \mu_W) \xrightarrow{d} N(\boldsymbol{0}_m, \boldsymbol{\Sigma}_W)$.

- **Asymptotic normality of $\hat{\beta}$:** Given the CLT and if HPs 1-5 of the MLG are satisfied, then the OLS estimator of β is asymptotically normal

$$\sqrt{T}(\widehat{\beta} - \beta) \simeq N(0, \sigma^2 Q^{-1})$$

3. Collinearity
Econometrics I

Collinearity refers to a problem that arises when independent variables are correlated.
The collinearity implies that multiple regression coefficient estimates can change erratically for small changes in model specification or changes in the data. In addition, a high degree of collinearity may cause problems when computing the inverse matrix of $(X'X)^{-1}$ necessary to calculate the regression coefficients.

Consider the following regression:

(1) $$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i, \quad u_i \sim N(0, \sigma^2)$$

and this auxiliary regression

(2) $$x_{i1} = \alpha_0 + \alpha_1 x_{i2} + u_i$$

Let R^2_x be the coefficient of determination of the second regression. How do you think its value will be? What does it entail?

3.collinearity
Econometrics I

There are **three degrees** of collinearity:

1) **Total absence of collinearity**. There is no correlation between the explanatory variables in the model.

2) **Presence of a certain degree of collinearity**. There is a high degree of linear correlation between some explanatory variables.

3) **Presence of perfect collinearity**. There are some explanatory variables that can be obtained from the linear combination of other explanatory variables, which implies that some explanatory variables are linearly dependent on each other. In this case, the estimation of the model is impossible due to the impossibility of inverting the X'X matrix.

3.collinearity
Econometrics I

If we analyze regression (1), we can write

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2(1 - R_x^2)}$$

the higher R^2_x is, the greater the variance. If R^2_x=0, then

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2}$$

- ❖ If we add variables in the model that are correlated, our estimates will be less accurate
- ❖ This clarify why we can't add variables that are linear combinations of the others. If the correlation between x_1 and x_2 was 1, then R^2_x= 1 and the variance of $\hat{\beta}_1$ would be infinite.
- ❖ VIF = $\frac{1}{1-R_x^2}$ is used to evaluate the possible existence of collinearity.

3.collinearity
Econometrics I

solutionscollinearity:

1. Add extra-sample information
2. Use estimators alternative to the OLS
3. Eliminate the non-significant variable.
   1. add one variable that is correlated with the others makes them less accurate estimates
   2. Should we remove it from the model?
      - If it is a non-relevant variable (the coefficient is statistically undistinguishable from zero) ➔ YES
      - But if it's a relevant variable (the coefficient is significant) ➔ NO. The reason is that if we remove a relevant variable we will have a bias in the estimators.

3.collinearity
Econometrics I

Example:

Given the following regression model, where x_1 and x_2 are relevant variables

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i \qquad (\beta_1 \neq 0,\ \beta_2 \neq 0).$$

We decide not to include x_2 and estimate model by OLS:

$$y_i = \beta_0 + \beta_1 x_{i1} + u_i \qquad \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)y_i}{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2}$$

$$\begin{aligned} \hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i)}{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2} \\ &= \beta_1 + \frac{\beta_2 \sum_{i=1}^{n}(x_{i1} - \bar{x}_1)x_{i2}}{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2} + \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)u_i}{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2} \end{aligned}$$

3.collinearity
Econometrics I

Then the OLS is a biased estimator $E(\beta_1) \neq \beta_1$

$$
\begin{aligned}
E(\hat{\beta}_1 | x_1, x_2) &= \beta_1 + \frac{\beta_2 \sum_{i=1}^{n}(x_{i1} - \bar{x}_1)x_{i2}}{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2} + \\
&\quad + \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)E(u_i | x_1, x_2)}{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2} \\
&= \beta_1 + \frac{\beta_2 \sum_{i=1}^{n}(x_{i1} - \bar{x}_1)x_{i2}}{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2}
\end{aligned}
$$