ECO375 Applied Econometrics I

Lecture Slide Notes

Tianyu Du

December 3, 2018

Updated version can be found on `www.tianyudu.com/notes`

Contents

4 Slide 4: Simple & Multiple Regression - Estimation

4.1 Regression Model

Assumption 4.1. Assuming the population follows

$$y = \beta_0 + \beta_1 x + u$$

and assume that x *causes* y.

4.2 OLS

$$\min_{\vec{\beta}} \sum_i (y_i - \hat{y}_i)^2$$

With FOC:

$$\sum_i (y_i - \hat{y}_i) = 0$$

$$\sum_i x_{ij}(y_i - \hat{y}_i) = 0, \ \forall j$$

Remark 4.1. Both $\hat{\beta}_0$ and $\hat{\beta}_j$ are functions of *random variables* and therefore themselves *random* with *sampling distribution*. And the estimated coefficients are random up to random sample chosen.

Proposition 4.1. Properties of OLS estimators

- **Unbiased** $\mathbb{E}[\hat{\beta}|X] = \beta$
- **Consistent** $\hat{\beta} \to \beta$ as $n \to \infty$
- **Efficient** min variance.

Definition 4.1. The **Simple Coefficient of Determination**

$$R^2 = \frac{SSE}{SST}$$

and $SS\underline{Total} = SS\underline{Explained} + SS\underline{Residual}$

$$\sum_i (y_i - \overline{y})^2 = \sum_i (\hat{y}_i - \overline{y})^2 + \sum_i (y_i - \hat{y}_i)^2$$

Proposition 4.2 (Logarithms)**.** Interpretation with logarithmic transformation.

- $\ln y = \alpha + \beta \ln y + u$: $\underline{x \text{ increases by } 1\%, \, y \text{ increases by } \beta\%}$.
- $\ln y = \alpha + \beta x + u$: $\underline{x \text{ increases by 1 unit}, \, y \text{ increases by } 100\beta\%}$.

- $y = \alpha + \beta \ln x + u$: x increases by 1%, y increases by 0.01β unit.

Assumption 4.2. (SLR) Simple regression model assumptions

1. Model is linear in parameter.
2. Random samples $\{(x_i, y_i)\}_{i=1}^n$.
3. Sample outcomes $\{x_i\}_{i=1}^n$ are not the same.[1]
4. $\mathbb{E}[u|x] = 0$ conditional on random sample x.
5. Error is homoskedastic. $Var(u|x) = \sigma^2$ for all x.

Benefits of MLR compared with SLR

- More accurate causal effect estimation.
- More flexible function forms.
- Could explicitly include more predictors so $\mathbb{E}(u|\mathbf{x}) = 0$ is easier to be satisfied. [2]
- MLR4 is less restrictive than SLR4.

Proposition 4.3. MLR OLS residual satisfies

$$\sum_i \hat{u}_i = 0$$

$$\sum_i x_{ji}\hat{u}_i = 0, \ \forall i \in \{1, 2, \dots, k\}$$

Proposition 4.4. MLR OLS estimators $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ pass through the average point.

$$\overline{y} = \hat{\beta}_0 + \hat{\beta}_1\overline{x}_1 + \cdots + \hat{\beta}_k\overline{x}_k$$

Proof.

$$\begin{aligned}
&\sum_i \hat{u}_i = 0 \\
\Longrightarrow &\sum_i \hat{y}_i - y_i = 0 \\
\Longrightarrow &\sum_i \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_k x_{ki} - y_i = 0 \\
\Longrightarrow &\ \hat{\beta}_0 + \hat{\beta}_1\overline{x}_1 + \cdots + \hat{\beta}_k\overline{x}_k = \overline{y}
\end{aligned}$$

■

[1] Otherwise, $SST = 0$.

[2] If we suspect some predictors may interact with certain component in u, moving the portion of u to the predictor set solves the problem.

4.3 Partialling Out

4.3.1 Steps

1. Regress x_1 on $x_2, x_3, \ldots, x_K$ and calculate the residual $\widetilde{r}_1$.

2. Regress y on $\widetilde{r}_1$ with simple regression and find the estimated coefficient $\hat{\lambda}_1$.

3. Then the multiple regression coefficient estimator $\hat{\beta}_1$ is

$$\hat{\beta}_1 = \hat{\lambda}_1 = \frac{\sum_i y_i \widetilde{r}_{1i}}{\sum_i (\widetilde{r}_{1i})^2}$$

Proof. By the first order condition of OLS,

$$\begin{aligned}
&\sum x_1 \hat{u} = 0 \\
\implies &\sum (\hat{x}_1 + \tilde{r}_1)\hat{u} = 0 \\
\implies &\sum \tilde{r}_1 \hat{u} = 0 \\
\implies &\sum \tilde{r}_1 (y - \hat{y}) = 0 \\
\implies &\sum \tilde{r}_1 (y - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_2 - \cdots - \hat{\beta}_k x_k) = 0 \\
\implies &\sum \tilde{r}_1 y = \hat{\beta}_1 \sum \tilde{r}_1 x_1 \\
\implies &\sum \tilde{r}_1 y = \hat{\beta}_1 \sum \tilde{r}_1 (\hat{x}_1 + \tilde{r}_1) = \hat{\beta}_1 \sum \tilde{r}_1{}^2 \\
\implies &\hat{\beta}_1 = \frac{\sum \tilde{r} y}{\sum \tilde{r}^2}
\end{aligned}$$

■

4.3.2 Interpretation

This OLS estimator only uses the unique variance of one independent variable. And the parts of variation correlated with other independent variables is partialled out.

Assumption 4.3. (MLR) Multiple Regression Assumptions

1. (MLR.1) The model is linear in parameters.

2. (MLR.2) Random sample from population $\{(x_{1i}, \ldots x_{ki}, y_i\}_{i=1}^n$.

3. (MLR.3) No perfect multicollinearity.

4. (MLR.4) Zero expected error conditional on population slice given by X.

$$\mathbb{E}(u|\mathbf{x}) \equiv \mathbb{E}(u|x_1, x_2, \ldots, x_k) = 0$$

5. (MLR.5) Homoskedastic error conditional on population slice given by X.

$$Var(u|\mathbf{x}) = \sigma^2$$

6. (MLR.6, *strict assumption*) Normally distributed error

$$u \sim \mathcal{N}(0, \sigma^2)$$

4.4 Omitted Variable Bias

Suppose population follows the *real model*

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + u_i \tag{1}$$

Consider the *alternative model*, and x_k is omitted, which is assumed to be relevant.

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_{k-1} x_{(k-1)i} + r_i \tag{2}$$

and use the partialling-out result on the second regression we have

$$\tilde{\beta}_1 = \frac{\sum_i \tilde{r}_{1i} y_i}{\sum_i (\tilde{r}_{1i})^2}$$

where

$$\tilde{r}_{1i} = x_{1i} - \tilde{\alpha}_0 - \tilde{\alpha}_2 x_{2i} - \cdots - \tilde{\alpha}_{k-1} x_{(k-1)i}$$

and

$$\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_k \frac{\sum_i (\tilde{r}_{1i} x_{ki})}{\sum_i (\tilde{r}_{1i})^2} \tag{3}$$

and take the expectation

$$\mathbb{E}(\tilde{\beta}_1 | X) = \beta_1 + \tilde{\delta}_1 \beta_k$$
$$Bias(\tilde{\beta}_1) = \tilde{\delta}_1 \beta_k$$

Conclusion the sign of bias depends on $Cov(x_1, x_k)$ and β_k.

Proof. TODO ■

5 Slide 5: Matrix Algebra for Regression Analysis

$$\mathbf{y} = \mathbf{A}\mathbf{x} \implies \frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{A} \tag{1}$$

Let $\alpha = \mathbf{y}'\mathbf{A}\mathbf{x}$, notice that $\alpha \in \mathbb{R}$, then

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{y}'\mathbf{A} \tag{2}$$
$$\frac{\partial \alpha}{\partial \mathbf{y}} = \mathbf{x}'\mathbf{A}' \tag{3}$$

Consider special case $\alpha = \mathbf{x}'\mathbf{A}\mathbf{x}$, then

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{x}'\mathbf{A} + \mathbf{x}'\mathbf{A}' \tag{4}$$

and if $\mathbf{A}$ is symmetric,

$$\frac{\partial \alpha}{\partial \mathbf{x}} = 2\mathbf{x}'\mathbf{A} \tag{5}$$

6 Slide 6: Multiple Regression in Matrix Algebra

6.1 The Model

Independent Variable Matrix

$$\mathbf{X} \in \mathbb{M}_{n\times(k+1)}(\mathbb{R})$$

where n is the number of observations and k is the number of features.

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2k} \\ \vdots & & & \\ 1 & x_{n1} & \dots & x_{nk} \end{bmatrix}_{n\times(k+1)}$$

Model

$$\mathbf{y} = \mathbf{X}\vec{\beta} + \mathbf{u}$$

First order condition for OLS

$$\frac{\partial(\mathbf{y} - \mathbf{X}\vec{\beta})^2}{\partial\vec{\beta}} = \mathbf{0} \iff \mathbf{X}'(\mathbf{y} - \mathbf{X}\vec{\beta}) = \mathbf{0}$$

Estimator

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Proof. From the first order condition for the OLS estimator

$$\begin{aligned} \mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\beta}) &= \mathbf{0} \\ \implies \mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\hat{\beta} &= \mathbf{0} \\ \implies \mathbf{X}'\mathbf{y} &= \mathbf{X}'\mathbf{X}\hat{\beta} \\ \implies \hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \end{aligned}$$

and

Remark 6.1. note that $(\mathbf{X}'\mathbf{X})$ is guaranteed to be invertible by assumption *no perfect multi-collinearity* and the implicit assumption that the number of features k is sufficiently greater than the number of observations n. i.e. $k >> n$.

■

Sum Squared Residual

$$SSR(\hat{\beta}) = ||\hat{u}||^2 = ||\mathbf{y} - \mathbf{X}\hat{\beta}||^2$$

6.2 Variance Matrix

Consider

$$\begin{aligned}\mathbf{z}_t &= [z_{1t}, z_{2t}, \dots z_{nt}]' \\ \mathbf{z}_s &= [z_{1s}, z_{2s}, \dots z_{ns}]'\end{aligned}$$

Notice that the variance and covariance are defined as

$$\begin{aligned}Var(\vec{z}_t) &= \mathbb{E}[(\vec{z}_t - \mathbb{E}[\vec{z}_t])^2] \\ Cov(\vec{z}_t, \vec{z}_s) &= \mathbb{E}[(\vec{z}_t - \mathbb{E}[\vec{z}_t])(\vec{z}_s - \mathbb{E}[\vec{z}_s])]\end{aligned}$$

The **variance matrix** of $\mathbf{z} = [z_1, z_2, \dots, z_n]$ is given by

$$\begin{aligned}Var(\mathbf{z}) &= \begin{bmatrix} Var(z_1) & Cov(z_1, z_2) & \dots & Cov(z_1, z_n) \\ Cov(z_2, z_1) & \dots & & \\ \vdots & & & \\ Cov(z_n, z_1) & \dots & \dots & Var(z_n) \end{bmatrix} \\ &= \begin{bmatrix} \mathbb{E}[(z_1 - \overline{z}_1)^2] & \mathbb{E}[(z_1 - \overline{z}_1)(z_2 - \overline{z}_2)] & \dots \\ \mathbb{E}[(z_2 - \overline{z}_2)(z_1 - \overline{z}_1)] & \dots & \\ \vdots & & \\ \mathbb{E}[(z_n - \overline{z}_n)(z_1 - \overline{z}_1)] & \dots & \mathbb{E}[(z_n - \overline{z}_n)^2] \end{bmatrix} \\ &= \mathbb{E}[(\mathbf{z} - \mathbb{E}[\mathbf{z}])_{n \times 1} \cdot (\mathbf{z} - \mathbb{E}[\mathbf{z}])'_{1 \times n}] \in \mathbb{M}_{n \times n}\end{aligned}$$

In the special case $\mathbb{E}[\mathbf{z}] = \mathbf{0}$, variance is reduced to

$$Var(\mathbf{z}|\mathbf{X}) = \mathbb{E}[\mathbf{z} \cdot \mathbf{z}'|\mathbf{X}]$$

Residual Since residual u_i are $i.i.d$ with variance σ^2, the variance matrix of $\mathbf{u}$ is

$$Var(\mathbf{u}|\mathbf{X}) = \mathbb{E}[\mathbf{u} \cdot \mathbf{u}'|\mathbf{X}] = \sigma^2 \mathbf{I}_n$$

Estimator If $\hat{\beta}$ is unbiased, $\mathbb{E}[\hat{\beta}|\mathbf{X}] = \vec{\beta}$, then

$$Var(\hat{\beta}|\mathbf{X}) = \mathbb{E}[(\hat{\beta} - \vec{\beta}) \cdot (\hat{\beta} - \vec{\beta})'|\mathbf{X}] \in \mathbb{M}_{(k+1) \times (k+1)}$$

7 Slide 7: Multiple Regression - Properties

7.1 Assumptions (MLRs) in Matrix Form

E.0. Random sample from population.

E.1. Linear in parameter

$$\mathbf{y} = \mathbf{X}\vec{\beta} + \mathbf{u}$$

E.2. No perfect multi-collinearity

$$rank(\mathbf{X}) = k + 1$$

E.3. Error has expected value of $\mathbf{0}$ conditional on $\mathbf{X}$ and $\mathbf{X}$ is orthogonal to residual $\mathbf{u}$.

$$\mathbb{E}[\mathbf{u}|\mathbf{X}] = \mathbf{0}$$

E.4. Error $\mathbf{u}$ is homoskedastic.

$$Var(\mathbf{u}|\mathbf{X}) = \sigma^2\mathbf{I}_n$$

E.5. Normally distributed error $\mathbf{u}$. Note that this assumption is relatively strong.

$$\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_n)$$

7.2 Properties of OLS Estimator

Theorem 7.1. Given *E.1. E.2. E.3.*, the OLS estimator $\hat{\beta}$ is an unbiased estimator for $\vec{\beta}$.

$$\mathbb{E}[\hat{\beta}|\mathbf{X}] = \vec{\beta}$$

Proof.

$$\begin{aligned}
\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\vec{\beta} + \mathbf{u}) \\
&= \vec{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}
\end{aligned}$$

Taking expectation conditional on $\mathbf{X}$ on both sides,

$$\mathbb{E}[\hat{\beta}|\mathbf{X}] = \vec{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{0} = \vec{\beta}$$

∎

Lemma 7.1. Suppose $\mathbf{A} \in \mathbb{M}_{m\times n}$ and $\mathbf{z} \in \mathbb{M}_{n\times 1}$ then

$$Var(\mathbf{A}\mathbf{z}) = \mathbf{A}Var(\mathbf{z})\mathbf{A}'$$

Theorem 7.2. Given $E.1 \sim E.4$

$$Var(\hat{\beta}|\mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2$$

Proof.

$$
\begin{aligned}
Var(\hat{\beta}|\mathbf{X}) &= Var((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}|\mathbf{X}) \\
&= Var((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\vec{\beta}+\mathbf{u})|\mathbf{X}) \\
&= Var(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}|\mathbf{X}) \\
&\text{By the lemma above,} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Var(\mathbf{u}|\mathbf{X})[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']' \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Var(\mathbf{u}|\mathbf{X})\mathbf{X}''(\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}_n\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}
$$

■

Theorem 7.3 (Gauss-Markov Theorem)**.** Given $E.1. \sim E.4.$, i.e.

1. Models is linear in parameters.
2. No perfect multi-collinearity presents.
3. Error has expected value of zero conditional on $\mathbf{X}$.
4. Homoskedastic.

the OLS estimator is the best linear unbiased estimator (BLUE).
(*The best* here refers to the OLS has the least variance among all estimators.)

7.3 Variance Inflation

Let $j \in \{1, 2, \dots, k\}$, then the variance of an individual estimator on particular feature j is

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{(1-R_j^2)SST_j}$$

where

$$SST_j = \sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2$$

and R_j^2 is the coefficient of determination while regressing x_j on all other features $x_i, \forall i \neq j$.

Definition 7.1. The **variance inflation** on estimator for feature j is

$$VIF_j = \frac{1}{1-R_j^2}$$

Remark 7.1 (Interpretation)**.** the standard error of estimator on a particular variable $(\hat{\beta}_j)$ is *inflated* by it's(x_j) relationship with other explanatory variables. If a predictor is highly correlated with other predictors, it's estimated coefficient will be inefficient (i.e. with high variance/uncertainty)

Solutions to high VIF

1. Drop the highly inflated explanatory variable.
2. Use ratio $\frac{x_i}{x_j}$ instead.
3. Ridge regression.

Remark 7.2 (Interpretation)**.** VIF highlights the importance of **not** including redundant predictors.

8 Slide 8: Multiple Regression - Inference

Hypothesis Testing on multiple regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots \beta_k x_{ik} + u_i$$

8.1 t-test for significance of individual predicator

Test statistic Given $MLR.1 \sim MLR.6$,
(requires $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$ so that t-statistic follows the t distribution),

$$t = \frac{\hat{\beta}_j - b}{s.e.(\hat{\beta}_j)} \sim t_{n-k-1}$$

where

$$H_0 : \beta_j = b$$
$$H_1 : \beta_j (\neq, >, <) b$$

8.2 t-test for comparing 2 coefficients

Test statistic

$$t = \frac{(\hat{\beta}_i - \hat{\beta}_j) - b}{s.e.(\hat{\beta}_i - \hat{\beta}_j)} \sim t_{n-k-1}$$

where

$$H_0 : \beta_i - \beta_j = b$$
$$H_1 : \beta_i - \beta_j (\neq, >, <) b$$

notice

$$s.e.(\hat{\beta}_i - \hat{\beta}_j) = \sqrt{Var(\hat{\beta}_i - \hat{\beta}_j)}$$
$$= \sqrt{Var(\hat{\beta}_i) + Var(\hat{\beta}_j) - 2Cov(\hat{\beta}_i, \hat{\beta}_j)}$$

8.3 Partial F-test for joint significance

$$H_0 : \beta_i = \beta_j = \beta_k = \cdots = 0$$
$$H_1 : \exists\ z \in \{i, j, k, \dots\}\ s.t.\ \beta_z \neq 0$$

Test significance by comparing the restricted and unrestricted models, see whether restricting the model by removing certain explanatory variables "significantly" hurts the fit of the model.

$$df = (q, n-k-1)$$

Test statistic Let SSR denote the regression residual,

$$F = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)} \sim F_{(q,n-k-1)}$$

or

$$F' = \frac{(R^2_{ur} - R^2_r)/q}{(1-R^2_{ur})/(n-k-1)} \sim F_{(q,n-k-1)}$$

8.4 Full F-test for the significance of the model

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$$
$$H_1 : \exists\ i \in \{1, 2, \dots, k\}\ s.t.\ \beta_i \neq 0$$

Remark 8.1. R^2 version only and substitute $R^2_r = 0$ (restricted model explains nothing), since SSR_r is undefined.

Test statistic

$$F = \frac{R^2_{ur}/k}{(1-R^2_{ur})/(n-k-1)} \sim F_{(k,n-k-1)}$$

8.5 F-test for general restrictions

$$H_0 : \beta_1 + \beta_2 = 1$$
$$H_1 : \neg H_0$$

Procedure

1. Substitute the restriction in H_0 into the original model to derive the restricted model.

2. Estimate both the original and restricted models and calculate their SSR.

Test hypothesis with $F - statistic$

$$F = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)}$$

where q denotes the number of restrictions (equations) in H_0.

Remark 8.2. Use the SSR version only of $F - statistic$ only since the SST for restricted and unrestricted models are different.

Remark 8.3. We only reject or failed to reject H_0, we never accept H_0 in a hypothesis test.

9 Slide 9: Multiple Regression - Further Issues

9.1 Data Scaling

9.1.1 Mutiplier

1. Enlarge x_j by factor a: $\hat{\beta}_j$ shrinks by a.
2. Enlarge y by factor a: **all** $\hat{\beta}_i$ enlarged by a.
3. Test statistic $t = \frac{\hat{\beta}}{s.e.(\hat{\beta})} = \frac{a\hat{\beta}}{s.e.(a\hat{\beta})}$ is unaffected.

9.1.2 Standardization

Standardized variable For j^{th} observation of explanatory variable x,

$$z_j = \frac{x_j - \overline{x}}{\sigma_x}$$

which satisfies

$$\mathbb{E}[z_j] = 0,\ Var(z_j) = 1$$

Properties Consider model and find the estimator of regressing standardized y on standardized x.

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik} + \hat{u}_i$$

Since OLS estimator passes through the mean,

$$\begin{gathered}
\overline{y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{x}_1 + \ldots \hat{\beta}_k \overline{x}_k \\
\implies (y_i - \overline{y}) = \hat{\beta}_1 (x_{i1} - \overline{x}_1) + \cdots + \hat{\beta}_k (x_{ik} - \overline{x}_k) + \hat{u}_i \\
\implies \frac{y_i - \overline{y}}{\sigma_y} = \frac{\hat{\beta}_1 \sigma_{x_1}}{\sigma_y} \frac{x_{i1} - \overline{x}_1}{\sigma_{x_1}} + \cdots + \frac{\hat{\beta}_k \sigma_{x_k}}{\sigma_y} \frac{x_{ik} - \overline{x}_k}{\sigma_{x_k}} + \frac{\hat{u}_i}{\sigma_y} \\
\implies b_j = \frac{\hat{\beta}_j \sigma_{x_j}}{\sigma_y}
\end{gathered}$$

Remark 9.1 (Interpretation)**.** x_j increases by 1 **std**, y increases by $b_j = \frac{\hat{\beta}_j \sigma_{x_j}}{\sigma_y}$ **std**, *ceteris paribus.*

9.2 Logarithmic Function

Exact interpretation of log transformation.

$$\ln(y_i) = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots \hat{\beta}_k x_{ik} + \hat{u}_i$$

Derive.

$$\begin{aligned}
\ln(y_2) - \ln(y_1) &= \hat{\beta}_j \Delta x_j \\
\implies \ln(\frac{y_2}{y_1}) &= \hat{\beta}_j \Delta x_j \\
\implies \frac{y_2}{y_1} &= exp(\hat{\beta}_j \Delta x_j) \\
\implies \frac{y_2 - y_1}{y_1} &= \frac{y_2}{y_1} - 1 \\
\implies \%\Delta y &= exp(\hat{\beta}_j \Delta x_j) - 1
\end{aligned}$$

■

9.3 Quadratics and Polynomials

Model

$$y_i = \sum_{p=0}^{k} \beta_p x_i^p + u_i$$

Remark 9.2. Consider the **interpretation** and **turning points**.

9.4 Interaction Effects

Consider model

$$y = \beta_0 + \beta_1 x + \beta_2 z + \beta_3 xz + u$$

then

$$\frac{\partial y}{\partial x} = \beta_1 + \beta_3 z$$

1. The effects of change of x on y depends on z.
2. Interpretation: *evaluate* $\frac{\partial y}{\partial x}$ at a z point that we are interested in.
3. Use *conventional testing* (t-test) to check if interaction term is significant.

9.5 Regression Selection and Adjusted R-square

The adjusted R-square, $\overline{R^2}$, incorporates a *penalty* for including more regressors (if insignificant).

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

Remark 9.3. $\overline{R^2}$ increases when adding new regressor(or a group of regressors) if and only if the t value (F) for the individual regression(group of regressors) is more than 1.

9.6 Causal Mechanism

9.7 Confidence Interval for Prediction

Consider a prediction

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots \hat{\beta}_k x_k$$

Evaluate at an arbitrary data point (not necessarily an observation in sample)

$$\mathbf{c} = (c_1, c_2, \dots, c_k)$$

Then the estimation of y at $\mathbf{c}$ is

$$\begin{aligned}
\theta_0 &= \mathbb{E}[y|x_1 = c_1, x_2 = c_2, \dots x_k = c_k] \\
&= \beta_0 + \beta_1 c_1 + \beta_2 c_2 + \dots + \beta_k c_k \\
\implies \beta_0 &= \theta_0 - \beta_1 c_1 - \beta_2 c_2 - \dots - \beta_k c_k
\end{aligned}$$

substitute back into the model

$$y = \theta_0 + \beta_1 (x_1 - c_1) + \beta_2 (x_2 - c_2) + \dots + \beta_k x_k + u$$

And the margin of error of confidence interval of prediction of y at $\mathbf{c}$ can be found by inspecting the intercept on above regression.

$$ME = t_{\frac{\alpha}{2}} \times s.e.(intercept)$$

The center of confidence interval can be found from

$$\hat{\theta}_0 = \hat{\beta}_0 + \hat{\beta}_1 c_1 + \dots + \hat{\beta}_k x_k$$

The α confidence interval is given by

$$\hat{\theta}_0 \pm ME$$

10 Slide 10: Multiple Regression - Qualitative Information

10.1 Binary predictors

Remark 10.1. With binary independent variables, $MLR.1 \sim MLR.6$ still holds, but the interpretations are different.

10.1.1 On Intercept

$$y = \delta_0 + \delta_1 male + \dots + u$$

Remark 10.2. To avoid perfect multi-collinearity, never include all categories.

10.1.2 On Slopes

$$y = \delta_0 + (\delta_1 + \delta_2 male) \times education + \cdots + u$$

10.1.3 F-test(Chow test)

Test whether the true coefficients in 2 linear regression models (e.g. for different gender groups) are equal.

1. Restricted model (SSR_r)

$$y = \beta_0 + \beta_1 x + u$$

2. Unrestricted model (SSR_{ur})

$$y = (\beta_0 + \delta_0 indicator) + (\beta_1 + \delta_1 indicator)x + u$$

3. Test whether the additional factors in coefficients (δ_0, δ_1) are significant. ($q = 2$ in this case)

$$F = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)}$$

10.2 Linear Probability Model

Qualitative binary dependent variable

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u,\ y \in \{0, 1\}$$

Interpretation the model above predicts the probability of $y = 1$.

Proof.

$$\begin{aligned}\mathbb{E}[y|\mathbf{x}] &= 0 \times Pr(y = 0|\mathbf{x}) + 1 \times Pr(y = 1|\mathbf{x}) \\ &= Pr(y = 1|\mathbf{x})\end{aligned}$$

∎

Remark 10.3. $\beta_j = \frac{\partial P(\mathbf{x})}{\partial x_j}$ is the **response probability**, and $\hat{P}(\mathbf{x})$ is the **predicted probability** of y to be 1.

Remark 10.4 (Out-of-range predictions)**.** Notice the prediction is not necessarily with the range of $[0, 1]$ for some extreme values of $\mathbf{x}$.

10.3 Heterskedasticity of LPM

Remark 10.5. For probability linear models, $MLR.5$(homoskedasticity) fails.

Proof.

$$
\begin{aligned}
y_i &= \beta_0 + \beta_1 x_{i1} + \dots \beta_k x_{ik} + u_i \\
&\text{For binary } y \\
Var(u) &= Var(y) = Pr(y=1)(1 - Pr(y=1)) \\
Var(u|\mathbf{x}) &= Var(y - \beta_0 - \beta_1 x_1 - \beta_2 x_2 - \dots - \beta_k x_k|\mathbf{x}) \\
&= Var(y|\mathbf{x}) \\
&= Pr(y=1|\mathbf{x})(1 - Pr(y=1|\mathbf{x})) \\
&= \mathbb{E}[y|\mathbf{x}](1 - \mathbb{E}[y|\mathbf{x}]) \\
&= (\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)(1 - \beta_0 - \beta_1 x_1 - \dots - \beta_k x_k) \\
&\neq \sigma_u^2
\end{aligned}
$$

■

11 Slide 11: Heteroskedasticity

Definition 11.1. Consider model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + u_i$$

the error of above model is heteroskedastic if for each sample point $\mathbf{x}_i \in \mathbb{R}^{k+1}$,

$$Var(u_i|\mathbf{x}_i) = \sigma_i^2$$

and σ_i^2 is not the same for all i.

Remark 11.1 (Consequence)**.** Without $MLR.5$, Gauss-Markov theorem does not hold and

1. OLS estimator is still linear and unbiased.
2. But **not** necessarily the best (variance is affected).

Proof. unbiasedness, in simple regression.

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sum_i (x_i - \overline{x})^2} \\
&= \frac{\sum_i (x_i - \overline{x})(\beta_0 + \beta_1 x_1 + u_i - \overline{y})}{\sum_i (x_i - \overline{x})^2} \\
&= \frac{\sum_i (x_i - \overline{x})(\beta_0 + \beta_1 x_1 + \beta_1 \overline{x} - \beta_1 \overline{x} + u_i - \overline{y})}{\sum_i (x_i - \overline{x})^2} \\
&= \frac{\sum_i \beta_1 (x_i - \overline{x})^2 + (x_i - \overline{x})(\beta_0 + \beta_1 \overline{x} - \overline{y} + u_i)}{\sum_i (x_i - \overline{x})^2} \\
&= \beta_1 + \frac{\sum_i (x_i - \overline{x})(0 + u_i)}{\sum_i (x_i - \overline{x})^2} \\
&= \beta_1 + \frac{\sum_i (x_i - \overline{x}) u_i}{\sum_i (x_i - \overline{x})^2}
\end{aligned}
$$

taking expectation conditional on $\mathbf{x}$ on both sides

$$\mathbb{E}[\hat{\beta}_1 | \mathbf{x}] = \beta_1$$

■

Proof. variance.

$$
\begin{aligned}
Var(\hat{\beta}_1 | \mathbf{x}) &= \mathbb{E}[(\hat{\beta} - \mathbb{E}[\hat{\beta}_1 | \mathbf{x}])^2 | \mathbf{x}] \\
&= \mathbb{E}[(\hat{\beta}_1 - \beta_1)^2 | \mathbf{x}] \\
&= \mathbb{E}[(\frac{\sum_i (x_i - \overline{x}) u_i}{\sum_i (x_i - \overline{x})^2})^2 | \mathbf{x}] \\
&= \frac{\sum_i (x_i - \overline{x}) \mathbb{E}[u_i | \mathbf{x}]}{\left(\sum_i (x_i - \overline{x})^2\right)^2} \\
&\neq \frac{\sigma^2}{SST_x}
\end{aligned}
$$

For multiple regressions

$$Var(\hat{\beta}_j | \mathbf{x}) = \frac{\sum_i \tilde{r}_{ij}^2 \sigma_i^2}{SSR_j^2} \neq \frac{\sigma^2}{SSR_j} = \frac{\sigma}{(1 - R_j^2) SST_j}$$

■

Remedies

1. Change variables so that the new model is homoskedastic.
2. Use robust standard errors.
3. Generalized least square (GLS).

11.1 Robust Standard Errors

Idea use $\hat{u}_i^2$ to estimate σ_i^2.
Note that

$$\begin{aligned} Var(u_i|\mathbf{x}) &= \mathbb{E}[(u_i - \mathbb{E}[u_i])^2] \\ &= \mathbb{E}[u_i^2|\mathbf{x}] + \mathbb{E}[u_i|\mathbf{x}]^2 \\ &= \mathbb{E}[u_i^2|\mathbf{x}] \end{aligned}$$

Consider model

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

OLS estimator is

$$\begin{aligned} \hat{\beta}_1 &= \beta_1 + \frac{\sum_i (x_i - \overline{x}) u_i}{\sum_i (x_i - \overline{x})^2} \\ Var(\hat{\beta}|\mathbf{x}) &= \frac{\sum_i (x_i - \overline{x})^2 \sigma_i^2}{\sum_i (x_i - \overline{x})^2} \\ \widehat{Var}(\hat{\beta}|\mathbf{x}) &= \frac{\sum_i (x_i - \overline{x})^2 \hat{u}_i^2}{\sum_i (x_i - \overline{x})^2} \end{aligned}$$

11.2 Test for Heteroskedasticity

11.2.1 General Principle

$$\begin{aligned} H_0 &: \mathbb{E}[u_i^2] = Var(u_i|\mathbf{x}) = \sigma^2 \text{ (Homoskedastic)} \\ H_1 &: \mathbb{E}[u_i^2] = Var(u_i|\mathbf{x}) = \sigma_i^2 \text{ (Heteroskedastic)} \end{aligned}$$

Methodology: specify the variance in alternative hypothesis to be a specific function of $\mathbf{x}$ or y.
Consider the model:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + u_i$$

And H_1 can be expressed as

$$H_1 : \mathbb{E}[u_i^2|\mathbf{x}] = \delta_0 + \delta_1 z_1 + \delta_2 z_2 + \cdots + \delta_p z_p$$

then run the proxy hypothesis testing

$$\begin{aligned} H_0' &: \delta_1 = \delta_2 = \cdots = \delta_p = 0, \delta_0 = \sigma^2 \\ H_1' &: \exists j \ s.t. \ \delta_j \neq 0 \end{aligned}$$

Note that the restricted model is homoskedastic.
Firstly run the original regression model and get residual $\hat{u}_i$.
Then test the proxy hypotheses with regression $\hat{u}_i^2$ on $z_1, z_2, \ldots, z_p$ using full F-test.

$$F = \frac{R_{\hat{u}^2}^2 / p}{(1 - R_{\hat{u}^2}^2)/(n-p-1)} \sim F_{(p,n-p-1)}$$
$$\text{and } nR_{\hat{u}^2}^2 \sim \mathcal{X}_p^2$$

11.2.2 Breusch-Pagan test

Use regressors x_i for z_i.
Auxiliary regression:

$$\hat{u}_i^2 = \delta_0 + \delta_1 x_1 + \dots \delta_k x_k$$
$$nR^2_{\hat{u}^2} \sim \mathcal{X}^2_k$$

11.2.3 White test version 1

Use polynomials of x_i for z_i.
Auxiliary regression: (for the case of 2 regressors)

$$\hat{u}_i^2 = \delta_0 + \delta_{i1} x_1 + \delta_2 x_{i2} + \delta_3 x_{i1}^2 + \delta_4 x_{i2}^2 + \delta_5 x_{i1} x_{i2} + \epsilon$$
$$nR^2_{\hat{u}^2} \sim \mathcal{X}^2_5$$
$$\text{or full F-test}$$

11.2.4 White test version 2

Use predicted response $\hat{y}$ (since its a linear combination of predictors) and its polynomial as z_i.
Auxiliary regression:

$$\hat{u}_i^2 = \delta_0 + \delta_1 \hat{y} + \delta_2 \hat{y}^2 + \epsilon$$

With hypotheses

$$H_0 : \delta_1 = \delta_2 = 0$$
$$H_1 : \delta_1 \neq 0 \vee \delta_2 \neq 0$$

$$nR^2_{\hat{u}^2} \sim \mathcal{X}^2_2$$
$$\text{or full F-test}$$

11.3 Generalized/Weighted Least Squared

Motivation when a regression model is suspicious for *heteroskedasticity* (i.e. MLR5 fails), Gauss-Markov theorem does no longer hold and OLS still unbiased and consistent but no longer the most efficient one. We wish to *transform* the original model, by multiplying by weights(p_i), to a homoskedastic model. And then run OLS on the transformed model to get linear estimations for coefficients, which are efficient. (Guaranteed by Gauss-Markov theorem)

11.3.1 GLS with Known Functional Form

Suppose (central assumption)

$$Var(u_i) = \mathbb{E}[u_i^2|\mathbf{X}] = h_i\sigma^2$$

for some known function h_i. Take weight function

$$p_i := \frac{1}{\sqrt{h_i}}$$

The **transformed** equation becomes

$$\begin{aligned} p_i y_i &= p_i\beta_0 + \beta_1 p_i x_{i1} + \cdots + \beta_k p_i x_{ik} + p_i u_i \\ \iff y_i/\sqrt{h_i} &= \beta_0/\sqrt{h_i} + \beta_1(x_{i1}/\sqrt{h_i}) + \cdots + \beta_k(x_{ik}/\sqrt{h_i}) + u_i/\sqrt{h_i} \\ \implies \mathbb{E}[(u_i/\sqrt{h_i})^2|\mathbf{X}] &= \frac{1}{h_i}h_i\sigma^2 = \sigma^2 \end{aligned}$$

which is homoskedastic.

Remark 11.2. In weighted least square with weight function p_i above, the variance of residual at a certain data cross-section is proportional to h_i. And $p_i \equiv \frac{1}{\sqrt{h_i}}$, that's, *observations with higher residual variance receive less weight.*

11.4 Feasible GLS

Suppose (central assumption)

$$Var(u_i|\mathbf{X}) = \mathbb{E}[u_i^2|\mathbf{X}] = \sigma^2 exp(\vec{\delta}\cdot\mathbf{x}_i)$$

for some constant σ.
Equivalently,

$$h(x) = exp(\delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k)$$

Remark 11.3. The *exponential* operator in our assumption guarantees the error variance is strictly positive.

To estimate the **variance**, we are going to model the squared residual u^2,

$$\begin{aligned} u_i^2 &= \sigma^2 exp(\vec{\delta}\cdot\mathbf{x}_i)v_i \\ \iff \ln(u_i^2) &= [\ln(\sigma^2)+\delta_0] + [\delta_1 x_{i1} + \cdots + \delta_k x_{ik}] + e_i \\ \alpha_0 &\equiv \ln(\sigma^2)+\delta_0, \quad e_i \equiv \ln(v_i) \end{aligned}$$

Procedures of FGLS

1. Run OLS regression on the original model, then estimate $\hat{u}_i^2$ and $\ln(\hat{u}_i^2)$.
2. Estimate model

$$\ln(\hat{u_i}^2) = [\ln(\sigma^2)+\delta_0] + [\delta_1 x_{i1} + \cdots + \delta_k x_{ik}] + e_i$$

3. Compute
$$\hat{h}_i := exp(\widehat{\ln(u_i^2)})$$
using result from above model. And compute
$$p_i := \frac{1}{\sqrt{\hat{h}_i}}$$

4. Transform the original model using weights p_i and estimate it using OLS. Note that the transformed model has **no** constant term. The constant is replaced with $(p_i\beta_0)$, which varies across observations.

12 Slide 12: Specification and Data Problems

A multiple regression model suffers from functional misspecification when it does not properly account for the relationship between the dependent and the observed explanatory variables.

12.1 Regression Specification Error Test (RESET)

12.1.1 RESET: Nested Alternatives

Adding nonlinear functions of the regressors into the model and test for their significance.
Consider model
$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u \tag{1}$$
If the original model satisfies MLR.4 ($\mathbb{E}[u|\mathbf{X}] = 0$), then **no** nonlinear functions of the independent variables should be significant when added to equation (1).

Procedures

1. Add polynomials in the OLS fitted values, $\hat{y}$, to equation (1). Typically squared and cubed terms are added.
$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \delta_1 \hat{y}^2 + \delta_2 \hat{y}^3 + u \tag{2}$$

2. Use F-test to test the joint significance with $H_0 : \delta_1 = \delta_2 = 0$. And a significant F suggests some sort of functional form problem.
$$F \sim \mathcal{F}_{(2,n-k-2)}$$

Remark 12.1. We will not be interested in the estimated parameters from (2); we only use this equation to test whether (1) has missed important nonlinearities.

Remark 12.2 (Nested Alternatives)**.** One model is **nested** in another if you can always obtain the first model by constraining some of the parameters of the second model.

Example 12.1. In above example, the original regression is *nested* in the expanded regression. We can recover the original regression by constraining $\delta_1 = \delta_2 = 0$ in the expanded model.

12.1.2 Non-nested Alternatives: RESET

Neither of the two models below is nested in the other one, we **cannot** use F-test.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \tag{3}$$
$$y = \beta_0 + \beta_1 \log(x_1) + \beta_2 \log(x_2) + u \tag{4}$$

Procedures

1. Construct a *comprehensive model* that contains each model as a special case and then to test the restrictions that led to each of the models.

$$y = \beta_0 + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 \log(x_1) + \gamma_4 \log(x_4) + u \tag{5}$$

2. Test competing specifications
   (a) (F) test for specification (4): $H_0 : \gamma_1 = \gamma_2 = 0$.
   (b) (F) test for specification (3): $H_0 : \gamma_3 = \gamma_4 = 0$.

12.1.3 Non-nested alternatives: Davidson-MacKinnon test

Let $\hat{y}_3$ and $\hat{y}_4$ denote the fitted values from (3) and (4) respectively.
If model (3) holds with $\mathbb{E}[u|x_1, x_2] = 0$, the fitted values from the other model, (4), should be insignificant when added to equation (3).

Procedures

1. Test for specification (3) with $H_0 : \theta_1 = 0,\ H_1 : \theta_1 \neq 0$.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \theta_1 \hat{y}_4 + u \tag{6}$$

2. Test for specification (4) with $H_0 : \theta_1 = 0,\ H_1 : \theta_1 \neq 0$.
   A significant t statistic (against a two-sided alternative) is a rejection of (4).

$$y = \beta_0 + \beta_1 \log(x_1) + \beta_2 \log(x_2) + \theta_1 \hat{y}_3 + u \tag{7}$$

Remark 12.3 (Porblems)**.**

1. In Davison-MacKinnon test, its possible for us to reject or accept both specifications.
   (a) If neither rejected, use adjusted R-square to choose one model.
   (b) If both rejected, find another alternative.

2. Note that a rejection of (3) does not mean (4) is the correct model.

3. The case when competing models have different dependent variables could be problematic. ($y = \dots$ against $\log(y) = \dots$)

12.2 Proxy Variables

12.2.1 Procedures

Consider the true model

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k^* + u \tag{8}$$

where x_k^* is unobserved.

Notation 12.1. In this text, we always use starred-variable, var^*, to denote the true (sometime unobservable) variable.

(1) Selecting proxy Choose an observed variable x_k is a **proxy** for x_k^* such that

$$x_k^* = \delta_0 + \delta_k x_k + v \tag{9}$$

Remark 12.4. Typically we want $\delta_k > 0$, and no restriction on δ_0.

(2) Plug-in the Proxy Direct replacement

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k(\delta_0 + \delta_k x_k + v) + u \tag{10}$$
$$= (\beta_0 + \beta_k \delta_0) + \beta_1 x_1 + \cdots + \beta_k \delta_k x_k + (u + \beta_k v) \tag{11}$$

Assumption 12.1. For a *consistent* estimator, we need to assume that

1. u is uncorrelated with $x_1, x_2, \dots, x_k^*, x_k$.

2. v is uncorrelated with $x_1, x_2, \dots, x_k$.

$$\implies \mathbb{E}[x_k^* | x_1, x_2, \dots, x_k] \tag{12}$$
$$= \mathbb{E}[\delta_0 + \delta_k x_k + v | x_1, x_2, \dots, x_k] = \delta_0 + \delta_k x_k \tag{13}$$

To guarantee MLR.4 holds for both true model and the model with proxy substitution.

Remark 12.5. Under above assumptions and regressing y on $x_1, x_2, \dots, x_k$, the OLS estimator for $(\beta_1, \beta_2, \dots, \beta_{k-1})$ is still consistent and unbiased.
But for intercept and k^{th} coefficient, we are effectively estimating $\beta_0 + \delta_0 \beta_k$ and $\delta_k \beta_k$.

12.2.2 Proxy Bias

If x_k^* is correlated with all $\{x_1, x_2, \ldots, x_k\}$ (collinearity), i.e.

$$x_k^* = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_k x_k + v_k$$

the for the coefficient of x_j in the original regression,

$$plim(\hat{\beta}_j) = \beta_j + \beta_k \delta_j$$

which means the estimation is still biased. In this case, using a proxy variable will not solve the omitted variable bias problem.

12.3 Measurement Error in an Explanatory Variable

Consider the model

$$y = \beta_0 + \beta_1 x_1^* + u$$

but we can only observe $x_1 = x_1^* + e_1$.

Assumption 12.2. Assuming **measurement error** satisfies

$$\mathbb{E}[e_1] = 0$$

and the regression model becomes if we regress y on the observed x_1.

$$y = \beta_0 + \beta_1 x_1 + (u - \beta_1 e_1) \tag{14}$$

Assumption 12.3. u is uncorrelated with both x_1 and x_1^*, i.e. x_1 does not affect y after x_1^* has been controlled for.

12.3.1 Case 1: $Cov(x_1, e_1) = 0$

Remark 12.6. Since $e_1 = x_1 + x_1^*$, if $Cov(x_1, e_1) = 0$ then $Cov(x_1^*, e_1) \neq 0$.

Remark 12.7.

$$\mathbb{E}[u - \beta_1 e_1] = \mathbb{E}[u] - \beta_1 \mathbb{E}[e_1] = 0$$

MLR.3 still holds and estimator $\hat{\beta}_1$ is still consistent.

Remark 12.8. Note that

$$Var(u - \beta_1 e_1) = \sigma_u^2 + \beta_1^2 \sigma_{e_1}^2$$

the variance of estimators is inflated unless $\beta_1 = 0$.

12.3.2 Case 2 $Cov(x_1^*, e_1) = 0$: Classical errors-in-variance(CEV)

Remark 12.9.

$$\begin{aligned}
Cov(x_1, e_1) &= \mathbb{E}[(x_1 - \overline{x}_1)(e_1 - \overline{e}_1)] \\
&= \mathbb{E}[x_1 e_1] \\
&= \mathbb{E}[(x_1^* + e_1)e_1] \\
&= \mathbb{E}[x_1^* e_1 + e_1^2] \\
&= 0 + \mathbb{E}[e_1^2] \\
&= \mathbb{E}[(e_1 - \overline{e}_1)^2] \\
&= \sigma_{e_1}^2 \neq 0
\end{aligned}$$

Thus the covariance between x_1 and x_1 is equal to the variance of the measurement error under CEV assumption.

Remark 12.10. From equation (11), the new residual is $(u - \beta_1 e_1)$ and

$$\begin{aligned}
Cov(x_1, u - \beta_1 e_1) &= \sum (x_1 - \overline{x}_1)(u - \beta_1 e_1) \\
&= \sum x_1 u - \beta_1 \sum x_1 e_1 \\
&= Cov(x_1, u) - \beta_1 \sum (x_1 - \overline{x}_1)(e_1 - 0) \\
&= 0 - \beta_1 Cov(x_1, e_1) \\
&= \sigma_{e_1}^2 \neq 0
\end{aligned}$$

this fails MLR.4 and the OLS regression of y on x_1 gives a biased and inconsistent estimator.

12.4 Measurement Error in Dependent Variable

Consider model

$$y^* = \mathbf{X}\vec{\beta} + u \tag{15}$$

and the actually observed y is $y = y^* + e_0$, with **measurement error** e_0. If we regress the observed y on explanatory variables, we are effectively estimating

$$y = \mathbf{X}\vec{\beta} + (u + e_0) \tag{16}$$

Remark 12.11. Assuming the measurement error in y is statistically independent of each explanatory variable, the OLS estimator from (12) is consistent and unbiased (Gauss-Markov Holds).

Remark 12.12. Note that we would now have higher residual variance $\sigma_u^2 + \sigma_{e_0}^2$ and the variance for OLS estimator is inflated

$$Var(\vec{\beta}) = (\sigma_u^2 + \sigma_{e_0}^2)(\mathbf{X}'\mathbf{X})^{-1}$$

13 Slide 13: Instrumental Variables

13.1 Endogeneity

Definition 13.1. If a predictor x_j is correlated with u for any reason, and MLR.4 is violated, then x_j is said to be an **endogenous** explanatory variable.

$$\mathbb{E}[u|\mathbf{x}] \neq 0$$

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u \tag{1}$$

Sources of Endogeneity

- Omitted variable bias.
- Sample selection bias.
- Simultaneity (bidirectional causality).
- Measurement error bias.

Remedies

- Control for confounding variables.[3]
- Instrumental variables or two stage least square.
- Differences in difference. (repeated cross-section data)
- Fixed effects. (panel data)

13.2 Instrumental Variables

The Problem For the simple regression model

$$y = \beta_0 + \beta x + u$$

estimator $\hat{\beta}$ would be biased if endogeneity presents ($Cov(x, u) \neq 0$).
Then OLS is actually estimating

$$\frac{\partial y}{\partial x} = \beta + \frac{\partial u}{\partial x}$$

instead of purely β, where $\frac{\partial u}{\partial x} \neq 0$ due to endogeneity.
We need a method to generate only exogenous variation in x, without changing u, and measure its impact on y via β only.

[3] A **confounding variable** is a variable that influences both the dependent variable and independent variable causing a spurious association.

Definition 13.2. An **instrument** z for predictor x is a variable the property that

1. (Exogeneity condition) uncorrelated with u.

$$Cov(z, u) = 0$$

2. (Relevance condition) correlated (either positively or negatively) with x.

$$Cov(z, x) \neq 0$$

Remark 13.1. There no perfect test for exogeneity condition and we have to argue it by appealing to economic theory. So we cannot prove exogeneity condition formally.

Remark 13.2. For the relevance condition, we can test it by testing the significance of π_1 in the regression below

$$x = \pi_0 + \pi_1 z + v$$

13.3 Implementation of IV: Method of Moments

Procedure

1. Identify β in terms of *population moments*.

2. Replace the population moments with the sample moments.[4]

13.3.1 In Simple Regression

Identification Consider the model with instrumental variable z for x,

$$y = \beta_0 + \beta_1 x + u$$

subtract both sides the corresponding expectations,

$$y - \mathbb{E}[y] = \beta_1(x - \mathbb{E}[x]) + (u - \mathbb{E}[u])$$

multiplying both sides by $(z - \mathbb{E}[z])$ and take expectation

$$\mathbb{E}[(y - \mathbb{E}[y])(z - \mathbb{E}[z])] = \beta_1 \mathbb{E}[(x - \mathbb{E}[x])(z - \mathbb{E}[z])] + \mathbb{E}[(u - \mathbb{E}[u])(z - \mathbb{E}[z])]$$
$$\implies Cov(y, z) = \beta_1 Cov(x, z) + Cov(u, z)$$
$$\text{By exogeneity condition and relevance condition}$$
$$Cov(x, z) \neq 0 \wedge Cov(z, u) = 0$$
$$\implies \beta_1 = \frac{Cov(y, z)}{Cov(x, z)}$$

[4]By **analogy principle**, such replacement will lead to a consistent estimator.

Replacement calculate the sample covariances between y, z and x, z and substitute into above expression, the **IV estimator** of β_1 is

$$\hat{\beta}_1 = \frac{\sum_i (y_i - \overline{y})(z_i - \overline{z})}{\sum_i (x_i - \overline{x})(z_i - \overline{z})}$$

and the **IV estimator** of β_0 is

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

Remark 13.3. When $z = x$ the IV estimator is equivalent to the OLS estimator. And the IV estimator is consistent even when MLR.4 does not hold.

13.3.2 Inference

Assuming

$$\mathbb{E}[u^2|z] = \sigma^2 = Var(u)$$

Then the variance of $\hat{\beta}_1$ is

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{n\sigma_x^2 \rho_{x,z}^2}$$

with sample analogs and $R^2_{x,z}$ from regression of x_i on z_i, the estimated variance is

$$\widehat{Var(\hat{\beta}_1)} = \frac{\hat{\sigma}^2}{SST_x R^2_{x,z}}$$

Note that the variance of OLS estimator is estimated to be

$$\widehat{Var(\hat{\beta}_1)} = \frac{\hat{\sigma}^2}{SST_x}$$

Therefore the IV estimator is always larger than OLS variance.
Note that as $z \to x$, $R^2_{x,z} \to 1$ and IV estimator is approaching and ultimately equivalent to the OLS estimator.

13.3.3 Properties

If z and x are weakly correlated (aka. **weak instrument**).

- IV estimators can have large standard errors. (small $R^2_{x,z}$)
- IV estimators can have large asymptotic bias if $Corr(z, u) \neq 0$ (since we cannot check exogeneity condition formally, so we cannot rule out this probability).

For IV estimator,

$$plim \hat{\beta}_{1,IV} = \beta_1 + \frac{Corr(z, u)\sigma_u}{Corr(z, x)\sigma_x}$$

compared with OLS estimator

$$plim \hat{\beta}_{1,OLS} = \beta_1 + Corr(x, u)\frac{\sigma_u}{\sigma_x}$$

Remark 13.4. The R^2 in IV estimation can be negative, and we should be careful about interpreting R^2 in IV estimation.

13.4 IV in Multiple Regression

Consider the multiple regression model on k predictors, where y_2 is endogenous. The **structural model** is given in (2) below.

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \cdots + \beta_k z_{k-1} + u_1 \tag{2}$$

Identification Let z_k be an instrumental variable for y_2 the exogenity condition can be expressed as

$$Cov(z_k, u_1) = 0$$

and assuming all other explanatory variables z_i are uncorrelated with u_1. Also assume the *zero-mean-error*,

$$Cov(z_i, u_1) = 0, \ \forall i \in \{1, 2, \ldots, k-1\}$$
$$\mathbb{E}[u_1] = 0$$

Above conditions can be re-written as

$$\mathbb{E}[z_i u_1] = 0, \ \forall i \in \{1, 2, \ldots, k\}$$
$$\mathbb{E}[u_1] = 0$$

Above $k+1$ equations identify $\beta_0, \beta_1, \ldots, \beta_k$.

Replacement Replacing u_1 with $\hat{u}_1$ from regression (2),

$$\sum_{i=1}^{n}(y_{i1} - \hat{\beta}_0 - \hat{\beta}_1 y_{i2} - \hat{\beta}_2 z_{i1} - \cdots - \hat{\beta}_k z_{k-1}) = 0$$
$$\sum_{i=1}^{n} z_{i1}(y_{i1} - \hat{\beta}_0 - \hat{\beta}_1 y_{i2} - \hat{\beta}_2 z_{i1} - \cdots - \hat{\beta}_k z_{k-1}) = 0$$
$$\sum_{i=1}^{n} z_{i2}(y_{i1} - \hat{\beta}_0 - \hat{\beta}_1 y_{i2} - \hat{\beta}_2 z_{i1} - \cdots - \hat{\beta}_k z_{k-1}) = 0$$
$$\vdots$$
$$\sum_{i=1}^{n} z_{ik-1}(y_{i1} - \hat{\beta}_0 - \hat{\beta}_1 y_{i2} - \hat{\beta}_2 z_{i1} - \cdots - \hat{\beta}_k z_{k-1}) = 0$$

And solving above $k+1$ equations and replacing give the IV estimations of $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$.
The relevance condition $Corr(y_2, z_k)$ can be verified using **reduced-form(auxiliary) equation** below with $H_0 : \pi_k = 0$ and $H_1 : \pi_k \neq 0$.

$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \cdots + \pi_k z_k + v_2$$

14 Slide 14: Two Stage Least Square

14.1 Procedure

Motivation Multiple good instrumental variables for the endogenous variable.
Structural Equation:

$$y = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1 \tag{1}$$

with **Reduced Form Equation**:

$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + v_2 \tag{2}$$

where at least one of $\pi_2, \pi_2 \neq 0$. (Relevance condition)

2SLS Procedures

1. **Stage 1** Run regression on REF and compute $\hat{y}_2$, which is a linear combination of z_1, z_2, z_3. So $\hat{y}_2 \perp u_1$ by exogeneity condition. Note that, $v_2 \not\perp u_1$.
$$\hat{y}_2 = \hat{\pi}_0 + \hat{\pi}_1 z_1 + \hat{\pi}_2 z_2 + \hat{\pi}_3 z_3$$

2. Check significance of z_2 and z_3 to verify relevance condition.

3. **Stage 2** Regress y_1 on $\hat{y}_2$ and z_1 to obtain $\hat{\beta}_{1,2SLS}$.

Remark 14.1. The first stage of 2SLS removes endogeneity of y_2 (dropped with v_2).

2SLS Procedures: general case

1. **Stage 1** For each included endogenous explanatory variables, construct its reduced form equation with instrumental variables (excluded exogenous) and included exogenous variables.

2. Check significance of every instrumental variables using t test and/or the joint significance of all instrumental variables used.

3. **Stage 2** Regress y all included exogenous variables and the estimated reduced form equations ($\hat{y}_j$) for all included endogenous variables.

Remark 14.2 (Number of IVs, the general case)**.** With k predictors in total, if m of them are endogenous, we need at least m excluded exogenous variables to run 2SLS.
Otherwise, in the second stage regression, we would have less explanatory variables than parameters to be estimated. (*perfect collinearity*)

	H_0	H_1
$\hat{\vec{\beta}}_{OLS}$	Consistent and Efficient	Inconsistent
$\hat{\vec{\beta}}_{2SLS}$	Consistent but less Efficient	Consistent

14.2 Equivalence between IV and 2SLS

On the simple regression

$$y = \beta_0 + \beta_1 x + u$$

and let z be the excluded exogenous variable used as the instrumental for x.
For simplicity, assume $\overline{x} = \overline{y} = \overline{z} = 0$.
Then IV estimator

$$\hat{\beta}_{1,IV} = \frac{Cov(z,y)}{Cov(z,x)} = \frac{\sum yz}{\sum xz}$$

And 2SLS estimator

$$\begin{aligned}
\hat{\beta}_{1,2SLS} &= \frac{\sum(\hat{x} - \overline{\hat{x}})(y - \overline{y})}{\sum(\hat{x} - \overline{\hat{x}})^2} \\
&= \frac{\sum \hat{x} y}{\sum \hat{x}^2} = \frac{\sum(\hat{\pi}_0 + \hat{\pi}_1 z) y}{\sum(\hat{\pi}_0 + \hat{\pi}_1 z)^2} \\
&= \frac{\sum \hat{\pi}_1 yz}{\sum \hat{\pi}_1^2 z^2} = \frac{1}{\hat{\pi}_1} \frac{\sum yz}{\sum z^2} \\
&= \frac{\sum z^2}{\sum zx} \frac{\sum yz}{\sum z^2} = \frac{\sum yz}{\sum xz} = \hat{\beta}_{1,IV}
\end{aligned}$$

14.3 Evaluating 2SLS

14.3.1 Regressor Endogeneity

OLS is BLUE, if OLS is consistent we should not use the relatively less efficient 2SLS.

Hausman's Test for OLS Consistency If H_0 if failed to be rejected use OLS as BLUE, if we reject H_0 then use 2SLS.

$$H_0 : plim\ \hat{\beta}_{OLS} = plim\ \hat{\beta}_{2SLS} = \vec{\beta}$$
$$H_1 : plim\ \hat{\beta}_{OLS} \neq \vec{\beta} \wedge plim\ \hat{\beta}_{2SLS} = \vec{\beta}$$

Take

$$d = \hat{\beta}_{OLS} - \hat{\beta}_{2SLS}$$

Under the Null Hypothesis, a normalized d statistic is distributed as a χ_g where g is the number of parameters in the test.

14.3.2 Instrument Relevance

Check the significance of instrumental variables in **reduced form equations** with t-test or F-test. If certain IV is not significant in reduced form equation, then do not use this IV.

Consider model

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \beta_3 z_2 + u \tag{3}$$

where y_2 is suspended to be endogenous and (z_3, z_4) are used as instrumental variables.

14.3.3 Instrument Exogeneity

Theoretically impossible to test.

- Solution (1): economic sense.
- Solution (2): over-confidence test (with z_3 and z_4 as instrumental variables)

  1. Assume z_3 is a valid instrumental variable, use z_3 as IV to recover $\hat{u}_1$.
  $$\hat{u}_1 = y_1 - \hat{\beta}_{0,IV} - \hat{\beta}_{1,IV} y_2 - \hat{\beta}_{2,IV} z_1 - \hat{\beta}_{3,IV} z_3$$
  2. Test if $Cov(z_4, \hat{u}_1) = 0$ to test the validity of z_4.
  $$\hat{u}_1 = \delta_0 + \delta_1 z_1 + \delta_2 z_2 + \delta_3 z_3 + \delta_4 z_4 + \epsilon$$
  with H_0 all insignificant (exogenous) and H_1 at least one of z_i is significant (endogenous). And under H_0,
  $$nR^2_{u,z} \sim \chi^2_q$$
  where q is the **degree of overconfidence**, which is the number of IV excluded from the main regression minus the number of endogenous variables.
  3. Use z_4 to recover $\hat{u}_1$ and test again.

15 Slide 15: Simultaneous Equations

Motivation Variables are *jointly determined.*

Example 15.1. Linear supply and demand framework.

$$\begin{cases} p = \beta_{11} + \beta_{12} q_s + \beta_{13} z_1 + u_1 \\ p = \beta_{21} + \beta_{22} q_d + \beta_{23} z_1 + u_2 \\ p_d = p_s \end{cases}$$

where z_1 and z_2 are exogenous variables. (aka **supply and demand shifters**)
q_s, q_d, p are endogenous variables.
u_1 and u_2 are structural errors.

15.1 Simultaneity Bias in OLS

Above simultaneous model can be written as reduced from below

$$\begin{cases} q_s = q_d = \frac{1}{\beta_{12}-\beta_{22}}(\beta_{11} - \beta_{21} + \beta_{13}z_1 - \beta_{23}z_2 + u_1 - u_2) \\ p = \frac{\beta_{12}}{\beta_{12}-\beta_{22}}(\beta_{11} - \beta_{21} + \beta_{13}z_1 - \beta_{23}z_2 + u_1 - u_2) + \beta_{11} + \beta_{13}z_1 + u_1 \end{cases} \tag{1}$$

$$\implies \begin{cases} q_s = q_d = \pi_{11} + \pi_{12}z_1 + \pi_{13}z_2 + v_1 \\ p = \pi_{21} + \pi_{22}z_1 + \pi_{23}z_2 + v_2 \end{cases} \tag{2}$$

From (1), obviously q and p are correlated with u_1, u_2.
And $v_1, v_2 \not\perp u_1, u_2$.

Notation 15.1. β: **structural form parameters**. π: **reduced form parameters**. (v_1, v_2): **reduced form errors**.
And note that $z_1, z_2 \perp v_1, v_2$.

If we use (2) and OLS to regress q and p based on exogenous variables z_1, z_2, we will get consistent reduced form parameters but not structural form parameters.

15.2 IV Estimator and 2SLS

Key From (1) or (2), we can show that p, q are correlated with z_1, z_2. (Relevance condition)
Also $z_1, z_2 \perp v_1, v_2$ implies exogeneity condition holds.
Use z_1 and z_2 as instrument for p, q.

Procedures

1. (Stage 1 OLS) Regress q on z_1, z_2.
2. Estimate $\hat{q}. = \hat{\pi}_{11} + \hat{\pi}_{12}z_1 + \hat{\pi}_{13}z_2$.
3. (Stage 2 OLS)
   (a) Regress p on $\hat{q}, z_1$ to obtain supply function.
   (b) Regress p on $\hat{q}, z_2$ to obtain demand function.

Special case consider the case

$$\begin{cases} q_s = \alpha_1 p + \beta_1 z_1 + u_1 \\ q_d = \alpha_2 p + u_2 \\ q_s = q_d \end{cases} \tag{3}$$

We cannot recover q_s since $\hat{p}$ would be a function of z_1 only and we would encounter perfect co-linearity when regress q_s on $\hat{p}, z_1$.
In general, variables that appear **only** in the demand function can be valid instrument to estimate supply, vice versa. (Otherwise, perfect multi-collinearity)

Exclusion Restriction z_1 omitted in demand and z_2 omitted in supply.

Rank Condition (Sufficiency, not covered) tells us when such exclusion restrictions are sufficient to estimate structural parameters and ensures unique solution for structural parameters.

Order Condition (Necessary condition for identification) At least as many *excluded exogenous variables (instrument)* are required as *included endogenous variables* in the structural equation. (Otherwise, perfect multi-collinearity)

Remark 15.1. In 2SLS, we are basically replacing *included endogenous variables* with linear combinations (OLS prediction) of *all exogenous variables.* If one equation is unidentified, the total number of exogenous variables after replacement would be less than number of parameters

Definition 15.1. (Identifications)
Over identified equation: more excluded exogenous variables than included endogenous variables.
Just identified same number of excluded exogenous variables and included endogenous variables.
Unidentified less excluded exogenous variables than included endogenous variables.

Remark 15.2. Only over-identified and just identified equations can be correlated estimated by 2SLS.

17 Slide 17: Intro to Time Series

Definition 17.1. A **time series**(or stochastic process) is a sequence of random variables

$$\{y_t\}, \quad t = 1, 2, \ldots, n$$

- Order matters.
- Only a single *realization* of "economic history" (a stochastic process).
- Model the statistics as if we could observe repeated realization of the entire sequence.

Example 17.1 (Static Phillips Curve)**.** captures *contemporaneous relationships.*

$$inflation_t = \beta_0 + \beta_1 unemployment_t + u_t$$

Example 17.2 (Expectation Augmented Phillips Curve)**.**

$$\Delta infl_t = infl_t - infl_{t-1} = \beta_0 + \beta_1 unemp_t + u_t$$

17.1 Random Walk

Definition 17.2. **Random walk** process.

$$y_t = y_{t-1} + e_t$$

where e_t follows *white noise* satisfying

1. $\mathbb{E}[e_t] = 0$
2. $Var(e_t) = \sigma_e^2$
3. $\mathbb{E}[e_t e_s] = 0,\ t \neq s$ (i.e. $Cov(e_t, e_s) = 0,\ t \neq s$)

Properties of Random Walker Process

1. $y_t = \sum_{i=1}^{t} e_i + y_0$
2. $Var(y_t) = Var(y_0) + t^2\sigma_e^2$
3. $\mathbb{E}[y_t y_{t+h}] = \mathbb{E}[y_t(y_t + e_{t+1} + \cdots + e_{t+h})] = \mathbb{E}[y_t^2]$
   (a) If $\mathbb{E}[y_0] = Var(y_0)$ (i.e. $y_0 = 0$), then $\mathbb{E}[y_t^2] = t^2\sigma_e^2$.

17.2 Trend

- If trend presents (i.e. t should be included as a regressor), then ignoring the trend would introduce omitted variable bias.
- (In practice) partialling out trends.
- **Note:** R^2 tends to be high when using trending data and the high R^2 may reflect the explanatory power of trend (t) but not the explanatory power of $\mathbf{x}_t$

17.3 Seasonality

Simple Solution : including dummy variables representing different sections in a complete season loop.

$$y_t = \alpha_0 + \alpha_1 t + \delta_1 Q_{t1} + \delta_2 Q_{t2} + \delta_3 Q_{t3} + \beta\mathbf{x} + u_t$$

where $\alpha_1 t$ captures the trend and $\delta_i Qti$ captures the seasonalities.

17.4 Serial Correlation

Definition 17.3.

$$\mathbb{E}[u_t u_s|\mathbf{X}] \neq 0, \text{ for } t \neq s$$

Remark 17.1. Similar to heteroskedasticity, $\mathbb{E}[u_t^2|\mathbf{X}] = \sigma_t^2 \neq \sigma^2$.

Consequences OLS will still be

1. Consistent
2. and unbiased

but

1. not the best (i.e. not the most efficient)
2. biased/inconsistent variance-covariance matrix and biased inferences.

17.5 Stationarity

Definition 17.4. A *stochastic process*

$$\{x_t | t = 1, 2, \dots\}$$

is **stationary** if the *joint distribution* of $\{x_{t1}, x_{t2}, \dots\}$ is the same as $\{x_{t1+h}, x_{t2+h}, \dots\}$. i.e.

1. x_t is identically distributed for all $t \in \mathbb{N}$,
2. and $corr(x_t, x_{t+1}) = corr(x_{t+h}, x_{t+h+1})$.

17.6 Weakly Dependent Time Series

Definition 17.5. A time series is **weakly dependent** if, loosely speaking, x_t and x_{t+h} are "almost independent" as n increases without bound.
Similar to the *random sample* assumption in MLR.

$$\lim_{h \to \infty} corr(x_t, x_{t+h}) = 0$$

Remark 17.2. To use OLS on time series, the series must be both

1. Stationary
2. **and** weakly dependent

18 Slide 18: Asymptotic Analysis

Definition 18.1. A sequence of random variable $\{Z_n\}$ is said to **converge in distribution** to a random variable Z if

$$\lim_{n \to \infty} F_n(x) = F(x)$$

for $x \in \mathbb{R}$ at which F is continuous, where F_n is the *cdf* of Z_n and F is the *cdf* of Z.

$$Z_n \xrightarrow{d} Z$$

and F is the **limit distribution** of $\{Z_n\}$.

Theorem 18.1 (Classical Central Limit Theorem)**.** Let $\{X_1, \ldots, X_n\}$ be a sequence of n iid random variable, with

1. $\mathbb{E}[X_i] = \mu < \infty$ and
2. $0 < Var(X_i) = \sigma^2 < \infty$

Then as $n \to \infty$, the distribution of $\overline{X}_n \equiv \frac{\sum_{i=1}^n X_i}{n}$ converges to the *Normal distribution* with mean μ and variance $\frac{\sigma^2}{n}$, i.e.

$$\overline{X}_n \xrightarrow{d} \mathcal{N}(\mu, \frac{\sigma^2}{n})$$

irrespective of the shape of the original distribution of X_i.

Corollary 18.1.

$$\sqrt{n}\frac{(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} \mathcal{N}(0, 1)$$

Definition 18.2. Let $\{a_n\}$ be a sequence of *deterministic* (i.e. non-random) real numbers. If

$$\forall \epsilon > 0 \; \exists n^* \in \mathbb{N} \; s.t. \; |a_n - a| < \epsilon \; \forall n > n^*$$

then a_n **deterministically converges** to a as $n \to \infty$.
Equivalent notations:

1. $P(|a_n - a| > \epsilon) = 0, \; \forall n > n^*$
2. $a_t \to a$
3. $\lim(a_n) = 0$

Definition 18.3. Let $\{Z_n\}$ be a sequence of *random* variables. If

$$\forall \epsilon, \delta > 0 \; \exists n^* \in \mathbb{N} \; s.t. \; \forall n > n^*, \; P(|Z_n - a| > \epsilon) < \delta$$

then Z_n **converges in probability** to a as $n \to \infty$.
Equivalent notations:

1. $\lim_{n \to \infty} P(|Z_n - a| > \epsilon) = 0, \; \forall \epsilon > 0$
2. $Z_n \xrightarrow{p} a$
3. $plim(Z_n) = a$

Theorem 18.2. convergence in probability $\implies$ convergence in distribution. i.e.

$$Z_n \xrightarrow{p} a \implies Z_n \xrightarrow{d} a$$

Proof. ∎

Theorem 18.3 (Law of Large Numbers)**.** Let $\{X_1, \ldots, X_n\}$ be a sequence of n independent and iid random variables, with $\mathbb{E}[X_i] = \mu < \infty$. Then

$$plim_{n\to\infty}(\overline{X}_n) = \mu$$

Theorem 18.4 (Continuous Mapping Theorem (Transformation Theorem) I)**.** If $T_n \xrightarrow{p} a$ and $U_n \xrightarrow{p} b$, then

$$(T_n + U_n) \xrightarrow{p} a + b$$
$$T_n U_n \xrightarrow{p} ab$$
$$\frac{T_n}{U_n} \xrightarrow{p} \frac{a}{b} \text{ if } b \neq 0$$

Theorem 18.5 (Continuous Mapping Theorem (Transformation Theorem) II)**.** If $Z_n \xrightarrow{d} Z$ and $U_n \xrightarrow{p} b$, where Z is *random variable*, then

$$Z_n + U_n \xrightarrow{d} Z + b$$
$$Z_n U_n \xrightarrow{d} Zb$$
$$\frac{Z_n}{U_n} \xrightarrow{d} \frac{Z}{b} \text{ if } P(U_n = 0) = 0 \wedge b \neq 0$$

18.1 Estimator Properties

Definition 18.4. $\hat{\theta}_n$ is a **consistent** estimator of θ if and only if

$$\hat{\theta}_n \xrightarrow{p} \theta$$

Definition 18.5. $\hat{\theta}_n$ is an **unbiased** estimator of θ if and only if

$$\mathbb{E}[\hat{\theta}_n] = \theta$$

Example 18.1 (Unbiased but inconsistent)**.** Let $\{Z_n\}$ be a sequence of random variable, and

$$\mathbb{E}[Z_n] = \mu$$
$$\lim_{n\to\infty} Var(Z_n) \neq 0$$

Example 18.2 (Biased but consistent)**.** Let $\{Z_n\}$ be a sequence of random variable, and

$$\mathbb{E}[Z_n] = Z_n + \frac{c}{n},\ c \in \mathbb{R}$$
$$plim(\hat{\theta}_n) = \theta$$

Z_n is biased for small samples but consistent.

Proposition 18.1. If $\hat{\theta}_n$ is an unbiased estimator for θ and $\lim_{n\to\infty} \hat{\theta}_n = 0$, then $\hat{\theta}_n$ is a consistent estimator for θ, i.e.

$$\mathbb{E}[\hat{\theta}_n] = \theta \wedge \lim_{n\to\infty} Var(\theta_n) = 0 \implies plim(\hat{\theta}_n) = \theta$$

18.2 OLS Consistency

Remark 18.1. If MLR.6 holds, then the sampling distribution of OLSEs follows *exact* normal distribution, even with finite sample size, $n < \infty$.

Assumption 18.1 (MLR.4')**.**

$$\begin{aligned}\mathbb{E}[u] &= 0 \\ Cov(x_j, u) &= 0,\ \forall j = 1, \ldots, k\end{aligned}$$

Proposition 18.2. Note that MLR.4 $\implies$ MLR.4', i.e. MLR.4 is an assumption stronger than MLR.4'.

Proof.

$$\begin{aligned}\mathbb{E}[u] &= \mathbb{E}[\mathbb{E}[u|\mathbf{x}]] \\ &= \mathbb{E}[0] \text{ by MLR.4} \\ &= 0 \\ Cov(x_j, u) &= \mathbb{E}[x_j u] - \mathbb{E}[x_j]\mathbb{E}[u] \\ &= \mathbb{E}[x_j u] \\ &= \mathbb{E}[x_j \mathbb{E}[u|\mathbf{x}]] = 0\end{aligned}$$

■

Theorem 18.6 (OLSE Consistency)**.** Given assumption MLR.1 - MLR.4', the OLSE $\hat{\beta}$ is **consistent** for $\vec{\beta}$.

Proof.

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}) \\ &= \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u} \\ &= \boldsymbol{\beta} + (\frac{1}{n}\boldsymbol{X}'\boldsymbol{X})^{-1}\frac{1}{n}\boldsymbol{X}'\boldsymbol{u} \\ &= \boldsymbol{\beta} + (\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_t\boldsymbol{x}_t')^{-1}(\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_t u_t)\end{aligned}$$

By law of large number,

$$\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_t\boldsymbol{x}_t' \xrightarrow{p} \mathbb{E}[\boldsymbol{x}\boldsymbol{x}'] \equiv \boldsymbol{A}$$

and

$$\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_t u_t \xrightarrow{p} \mathbb{E}[\boldsymbol{x}u] = \boldsymbol{0}$$

So by continuous mapping theorem,

$$\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}\boldsymbol{A}^{-1}\mathbf{0} = \boldsymbol{\beta}$$

■

Remark 18.2. To prove the **unbiasedness**, we need MLR.4.

Remark 18.3 (On Omitted Variable Bias)**.** Omitted variable bias also violates MLR.4'. So omitted variable bias is called an **inconsistency bias** or **asymptotic bias** in $\hat{\boldsymbol{\beta}}$.

Example 18.3.

$$\tilde{\beta}_1 \xrightarrow{p} \beta_1 + \beta_2 \frac{Cov(x_1, x_2)}{Var(x_1)}$$

18.3 Asymptotic Normality

Theorem 18.7. Given assumptions MLR.1-MLR.5 or E.1-E.4, the OLSE, $\boldsymbol{\beta}$ is **asymptotically normal**, and

$$\hat{\boldsymbol{\beta}} \xrightarrow{p} \mathcal{N}(\boldsymbol{\beta}, \frac{\sigma^2 \boldsymbol{A}^{-1}}{n})$$

Proof. From the results above in consistency proof,

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{x}_t\boldsymbol{x}_t')^{-1}(\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{x}_t u_t)$$

By law of large number,

$$(\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{x}_t\boldsymbol{x}_t')^{-1} \xrightarrow{p} \boldsymbol{A}^{-1}$$

and let $i \neq j \in \{1, 2, \ldots n\}$, investigate

$$\begin{aligned} Cov(\boldsymbol{x}_i u_i, \boldsymbol{x}_j u_j) &= \mathbb{E}[\boldsymbol{x}_i u_i \boldsymbol{x}_j u_j] \\ &= \mathbb{E}[\boldsymbol{x}_i \boldsymbol{x}_j \mathbb{E}[u_i u_j | \boldsymbol{x}_i \boldsymbol{x}_j]] \\ &= \mathbb{E}[0] = 0 \end{aligned}$$

and for variance of $\boldsymbol{x}_t u_t$ for an arbitrary t,

$$\begin{aligned} Var(\boldsymbol{x}_t u_t) &= \mathbb{E}[\boldsymbol{x}_t \boldsymbol{x}_t' u_t^2] \\ &= \mathbb{E}[\mathbb{E}[\boldsymbol{x}_t \boldsymbol{x}_t' u_t^2 | \boldsymbol{x}_t]] \\ &= \mathbb{E}[\mathbb{E}[u_t^2 | \boldsymbol{x}_t] \boldsymbol{x}_t \boldsymbol{x}_t'] \\ &= \mathbb{E}[\sigma^2 \boldsymbol{x}_t \boldsymbol{x}_t'] \\ &= \sigma^2 \mathbb{E}[\boldsymbol{x}_t \boldsymbol{x}_t'] \\ &\equiv \sigma^2 \boldsymbol{A} \end{aligned}$$

Therefore, all $\boldsymbol{x}u$ are independent and we know $\mathbb{E}[\boldsymbol{x}u] = \mathbf{0}$. Therefore, by central limit theorem,

$$\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{x}_t u_t \overset{a}{\sim} \mathcal{N}(\mathbf{0}, \frac{\sigma^2 \boldsymbol{A}}{n})$$

By continuous mapping theorem, $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ is also normally distributed, and has variance

$$\begin{aligned} Var(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \boldsymbol{A}^{-1} Var(\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{x}_t u_t)\boldsymbol{A}^{-1} \\ &= \frac{\sigma^2 \boldsymbol{A}^{-1}}{n} \end{aligned}$$

Therefore

$$\begin{aligned} &\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \xrightarrow{p} \mathcal{N}(\mathbf{0}, \frac{\sigma^2 \boldsymbol{A}^{-1}}{n}) \\ &\implies \hat{\boldsymbol{\beta}} \xrightarrow{p} \mathcal{N}(\boldsymbol{\beta}, \frac{\sigma^2 \boldsymbol{A}^{-1}}{n}) \end{aligned}$$

$\blacksquare$