ECO375 Applied Econometrics I

Lecture Slide Notes

Tianyu Du

October 29, 2018

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

Contents

1 Slide 4: Simple & Multiple Regression - Estimation 3

1.1 Regression Model 3

1.2 OLS 3

1.3 Partialling Out 4

1.3.1 Steps 4

1.3.2 Interpretation 5

1.4 Omitted Variable Bias 5

2 Slide 5: Matrix Algebra for Regression Analysis 6

3 Slide 6: Multiple Regression in Matrix Algebra 6

3.1 The Model 6

3.2 Variance Matrix 7

4 Slide 7: Multiple Regression - Properties 8

4.1 Assumptions (MLRs) in Matrix Form 8

4.2 Properties of OLS Estimator 8

4.3 Variance Inflation 9

5 Slide 8: Multiple Regression - Inference 10

5.1 t-test for significance of individual predicator 10

5.2 t-test for comparing 2 coefficients 10

5.3 Partial F-test for joint significance 11

5.4 Full F-test for the significance of the model 11

5.5 F-test for general restrictions 11

6 Slide 9: Multiple Regression - Further Issues 11
6.1 Data Scaling 11
6.1.1 Mutiplier 11
6.1.2 Standardization 12
6.2 Logarithmic Function 12
6.3 Quadratics and Polynomials 13
6.4 Interaction Effects 13
6.5 Regression Selection and Adjusted R-square 13
6.6 Causal Mechanism 13
6.7 Confidence Interval for Prediction 13

7 Slide 10: Multiple Regression - Qualitative Information 14
7.1 Binary predictors 14
7.1.1 On Intercept 14
7.1.2 On Slopes 14
7.1.3 F-test(Chow test) 14
7.2 Linear Probability Model 15
7.3 Heterskedasticity of LPM 15

8 Slide 11: Heteroskedasticity 15
8.1 Robust Standard Errors 17
8.2 Test for Heteroskedasticity 17
8.2.1 General Principle 17
8.2.2 Breusch-Pagan test 18
8.2.3 White test version 1 18
8.2.4 White test version 2 18

9 Slide 12: Specification and Data Problems 19
9.1 Regression Specification Error Test (RESET) 19
9.1.1 RESET: Nested Alternatives 19
9.1.2 RESET: Non-nested Alternatives 19
9.2 Davidson-MacKinnon test: non-nested alternatives 19
9.3 Proxy Variables 20
9.3.1 Procedures 20
9.3.2 Proxy Bias 20
9.4 Measurement Error in an Explanatory Variable 21
9.4.1 Case 1 $Cov(x_1, e_1) = 0$ 21
9.4.2 Case 2 $Cov(x_1^{true}, e_1) = 0$: Classical errors-in-variance(CEV) 21
9.5 Measurement Error in Dependent Variable 21

10 Slide 13: Instrumental Variables 22
10.1 Endogeneity 22
10.2 Instrumental Variables 22
10.3 Implementation of IV: Method of Moments 23
10.3.1 In Simple Regression 23
10.3.2 Inference 24

10.3.3 Properties 24
10.4 IV in Multiple Regression 25

1 Slide 4: Simple & Multiple Regression - Estimation

1.1 Regression Model

Assumption 1.1. Assuming the population follows

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

and assume that x *causes* y.

1.2 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 1.1. 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.

Property 1.1. Properties of OLS estimators

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

Definition 1.1. 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 1.1 (Logarithms)**.** Interpretation with logarithmic transformation.

- $\ln y = \alpha + \beta \ln y + u$: <u>$x$ increases by 1%, y increases by β%</u>.
- $\ln y = \alpha + \beta x + u$: <u>$x$ increases by 1 unit, y increases by 100β%</u>.
- $y = \alpha + \beta \ln x + u$: <u>$x$ increases by 1%, y increases by 0.01β unit.</u>

Assumption 1.2. Simple regression model assumptions

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

Benefits of MLR compared with SLR

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

Property 1.2. MLR OLS residual satisfies

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

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

Property 1.3. 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. ■

1.3 Partialling Out

1.3.1 Steps

1. Regress x_1 on $x_2, x_3, \dots, x_K$ 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. ■

1.3.2 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 1.3. Multiple Regression Assumptions

1. (MLR1) The model is linear in parameters.
2. (MLR2) Random sample from population $\{(x_{1i}, \dots x_{ki}, y_i\}_{i=1}^n$.
3. (MLR3) No perfect multicollinearity.
4. (MLR4) Zero expected error conditional on population slice given by X.
$$\mathbb{E}(u|X) = \mathbb{E}(u|x_1, x_2, \dots, x_k) = 0$$
5. (MLR5) Homoskedastic error conditional on population slice given by X.
$$Var(u|X) = \sigma^2$$
6. (MLR6, *strict assumption*) Normally distributed error
$$u \sim \mathcal{N}(0, \sigma^2)$$

1.4 Omitted Variable Bias

Suppose population follows the *real model*

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

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

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

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

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

where $\tilde{r_{1i}} = x_{1i} - \tilde{\alpha}_0 - \tilde{\alpha}_2 x_{2i} - \dots - \tilde{\alpha}_{k-1} x_{(k-1)i}$

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

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 ■

2 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,

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

3 Slide 6: Multiple Regression in Matrix Algebra

3.1 The Model

Predictor

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

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

$$\mathbf{X}'\hat{u} = \mathbf{0} \in \mathbb{R}^{k+1}$$
$$\iff \mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\beta}) = \mathbf{0} \in \mathbb{R}^{k+1}$$

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 note that $(\mathbf{X}'\mathbf{X})$ is guaranteed to be invertible by assumption *no perfect multi-collinearity.* ■

Sum Squared Residual

$$SSR(\hat{\beta}) = \hat{u}' \cdot \hat{u} = (\mathbf{y} - \mathbf{X}\hat{\beta})' \cdot (\mathbf{y} - \mathbf{X}\hat{\beta})$$

3.2 Variance Matrix

Consider

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

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 $\mathbb{E}[\vec{z}] = \vec{0}$, variance is reduced to

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

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

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

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

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

4 Slide 7: Multiple Regression - Properties

4.1 Assumptions (MLRs) in Matrix Form

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 $\mathbf{X}$.

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

E.4. Error $\mathbf{u}$ is *homoscedastic.*

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

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

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

4.2 Properties of OLS Estimator

Theorem 4.1. 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 4.1. Suppose $\mathbf{A} \in \mathbb{M}_{m\times n}$ and $\mathbf{z} \in \mathbb{M}_{n\times 1}$ then

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

Theorem 4.2. Given $E.1 \sim E.4$

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

Proof.

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

■

Theorem 4.3 (Gause-Markov)**.** Given $E.1. \sim E.4.$, the OLS estimator is the best linear unbiased estimator(BLUE).
(*The best* here means the OLS has the least variance among all estimators.)

4.3 Variance Inflation

Let $j \in \{1, 2, \ldots, k\}$, 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 4.1. The **variance inflation** on estimator for feature j is

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

Remark 4.1 (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.

Solutions to high VIF

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

Remark 4.2. VIF highlights the importantce of **not** including redundant predictors.

5 Slide 8: Multiple Regression - Inference

Hypothesis Testing on multiple regression model

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

5.1 t-test for significance of individual predicator

Test statistic Given $MLR.1 \sim MLR.6$ (need $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$),

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

where

$$\begin{aligned} H_0 &: \beta_j = b \\ H_1 &: \beta_j (\neq, >, <) b \end{aligned}$$

5.2 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

$$\begin{aligned} H_0 &: \beta_i - \beta_j = b \\ H_1 &: \beta_i - \beta_j (\neq, >, <) b \end{aligned}$$

notice

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

5.3 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

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

or

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

5.4 Full F-test for the significance of the model

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$$
$$H_1 : \exists\ i \in \{1, 2, \dots, 3\}\ s.t.\ \beta_i \neq 0$$

Remark 5.1. R^2 version only and substitute $R^2_r = 0$, since SSR_r is undefined.

Test statistic

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

5.5 F-test for general restrictions

Remark 5.2. Use the SSR version of $Fstatistic$ only since the SST for restricted and unrestricted models are different.

Remark 5.3. We only reject or failed to reject H_0, we never accept H_0 in a hypothesis test.

6 Slide 9: Multiple Regression - Further Issues

6.1 Data Scaling

6.1.1 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.

6.1.2 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,

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

Remark 6.1 (Interpretation)**.** x_j increases by 1 **std**, y increases by $b_j = \frac{\hat{\beta}_j \sigma_{x_j}}{\sigma_y}$ **std**, *ceteris paribus.*

6.2 Logarithmic Function

Exact interpretation of log transformation.

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

Derive.

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

■

6.3 Quadratics and Polynomials

Model

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

Remark 6.2. Consider the **interpretation** and **turning points**.

6.4 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.

6.5 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 6.3. $\overline{R^2}$ 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.

6.6 Causal Mechanism

6.7 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

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

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

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

The α confidence interval is given by

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

7 Slide 10: Multiple Regression - Qualitative Information

7.1 Binary predictors

Remark 7.1. With binary independent variables, $MLR.1 \sim MLR.6$ still holds, but the interpretations are different.

7.1.1 On Intercept

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

Remark 7.2. To avoid perfect multi-collinearity, never include all categories.

7.1.2 On Slopes

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

7.1.3 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)

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

7.2 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 7.3. $\beta_j = \frac{\partial P(\mathbf{x})}{\partial x_j}$ is the **response probability**, and $\hat{P}(\mathbf{x})$ is the **predicted probability** of y to be 1.

Remark 7.4 (Out-of-range predictions)**.** Notice the prediction is not necessarily with the range of $[0, 1]$ for some extreme values of $\mathbf{x}$.

7.3 Heterskedasticity of LPM

Remark 7.5. For probability linear models, $MLR.5$(homoskedasticity) fails.

Proof.

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots \beta_k x_{ik} + u_i$$

$$\text{For binary } y$$

$$Var(u) = Var(y) = Pr(y = 1)(1 - Pr(y = 1))$$

$$\begin{aligned}Var(u|\mathbf{x}) &= Var(y - \beta_0 - \beta_1 x_1 - \beta_2 x_2 - \cdots - \beta_k x_k|\mathbf{x}) \\ &= Var(y|\mathbf{x}) \\ &= Pr(y = 1|\mathbf{x})(1 - Pr(y = 1|\mathbf{x})) \\ &= \mathbb{E}[y|\mathbf{x}](1 - \mathbb{E}[y|\mathbf{x}]) \\ &= (\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)(1 - \beta_0 - \beta_1 x_1 - \cdots - \beta_k x_k) \\ &\neq \sigma_u^2\end{aligned}$$

■

8 Slide 11: Heteroskedasticity

Definition 8.1. Consider model

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

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

1. OLS estimator is still <u>linear</u> and <u>unbiased</u>.
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).

8.1 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

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

8.2 Test for Heteroskedasticity

8.2.1 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

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

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.

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

8.2.2 Breusch-Pagan test

Use regressors x_i for z_i.
Auxiliary regression:

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

8.2.3 White test version 1

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

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

8.2.4 White test version 2

Use <u>predicted</u> 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

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

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

9 Slide 12: Specification and Data Problems

9.1 Regression Specification Error Test (RESET)

9.1.1 RESET: Nested Alternatives

Adding nonlinear functions of the regressors into the model and test for their significance.
Consider model

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

1) Run above regression, estimate $\hat{y}$.
2) And add non-linear transformations of the **estimation** to construct an **expanded regression model**.

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \delta_1 \hat{y}^2 + \delta_2 \hat{y}^3 + u$$

3) Use F-test to test the joint significance with $H_0 : \delta_1 = \delta_2 = 0$.

Remark 9.1 (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 9.1. 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.

9.1.2 RESET: Non-nested Alternatives

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

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \tag{1}$$
$$y = \beta_0 + \beta_1 \log(x_1) + \beta_2 \log(x_2) + u \tag{2}$$

1) Expanded regression:

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

2-i) (F) test for specification (1): $H_0 : \gamma_1 = \gamma_2 = 0$.
2-ii) (F) test for specification (2): $H_0 : \gamma_3 = \gamma_4 = 0$.

9.2 Davidson-MacKinnon test: non-nested alternatives

Let $\hat{y}_1$ and $\hat{y}_2$ denote the fitted values from (1) and (2) respectively.

1-i) Test for specification (1) with $H_0 : \theta_1 = 0$. (If (1) is corrected, $\hat{y}_2$ should be insignificant.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \theta_1 \hat{y}_2 + u \tag{3}$$

1-ii) Test for specification (2) with $H_0 : \theta_1 = 0$:

$$y = \beta_0 + \beta_1 \log(x_1) + \beta_2 \log(x_2) + \theta_1 \hat{y}_1 + u \tag{4}$$

Remark 9.2. In Davison-MacKinnon test, its possible for us to reject or accept both specifications.

9.3 Proxy Variables

9.3.1 Procedures

For the original model

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k^{uob} + u \tag{5}$$

where x_k^{uob} is unobserved.

Select Choose an observed variable x_k is a **proxy** for x_k^{uob} such that

$$x_k^{uob} = \delta_0 + \delta_k x_k + v_3,\ \ \delta_k \neq 0 \tag{6}$$

Plug-in plug (6) into (5)

$$y = (\beta_0 + \beta_k \delta_0) + \beta_1 x_1 + \cdots + \beta_k \delta_k x_k + (u + \beta_k v) \tag{7}$$

Assumption 9.1. For a consistent estimator, we need to assume that

1. u is uncorrelated with $x_1, x_2, \ldots, x_k^{uob}, x_k$.
2. v is uncorrelated with $x_1, x_2, \ldots, x_k$.

$$\mathbb{E}[x_k^{uob} | x_1, x_2, \ldots, x_k] = \mathbb{E}[\delta_0 + \delta_k x_k + v | x_1, x_2, \ldots, x_k] = \delta_0 + \delta_k x_k$$

Remark 9.3. Under above assumptions, the OLS estimator for $(\beta_1, \beta_2, \ldots \beta_k \delta_k)$ is still consistent and unbiased.

9.3.2 Proxy Bias

If x_k^{uob} is correlated with all $\{x_1, x_2, \ldots, x_k\}$, i.e.

$$x_k^{uob} = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \cdots + \delta_k x_k + v_k$$

the for the coefficient of x_j in the original regression, we are effective estimating

$$\beta_j + \beta_k \delta_j$$

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

9.4 Measurement Error in an Explanatory Variable

Consider the model

$$y = \beta_0 + \beta_1 x_1^{true} + u$$

but we can only observe $x_1 = x_1^{true} + e_1$. Assuming **measurement error** satisfies

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

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

$$y = \beta_0 + \beta_1 x_1 + (u - \beta_1 e_1)$$

Assumption 9.2. u is uncorrelated with both x_1 and x_1^{true}.

9.4.1 Case 1 $Cov(x_1, e_1) = 0$

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

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

9.4.2 Case 2 $Cov(x_1^{true}, e_1) = 0$: Classical errors-in-variance(CEV)

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

Then the estimator $\hat{\beta}_1$ is biased and inconsistent.

9.5 Measurement Error in Dependent Variable

Consider model

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

and the actually observed y is $y = y^{true} + e_0$. If we regress the observed y on explanatory variables, we are effectively estimating

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

note that we would now have higher residual variance $\sigma^2 + \sigma_{e_0}^2$ and the variance for OLS estimator is inflated

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

If $\mathbb{E}[e_1|\mathbf{X}] = 0$ then $\hat{\beta}$ is still unbiased, otherwise in trouble.

10 Slide 13: Instrumental Variables

10.1 Endogeneity

Definition 10.1. 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.[1]
- Instrumental variables or two stage least square.
- Differences in difference. (repeated cross-section data)
- Fixed effects. (panel data)

10.2 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.

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

Definition 10.2. 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 10.1. 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.

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

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

10.3 Implementation of IV: Method of Moments

Procedure

1. Identify β in terms of *population moments*.
2. Replace the population moments with the sample moments.[2]

10.3.1 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

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

[2]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 10.2. 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.

10.3.2 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 $R_{x,z}^2$ from regression of x_i on z_i, the estimated variance is

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

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 $z \to x$, $R_{x,z}^2 \to 1$ and IV estimator is approaching and ultimately equivalent to the OLS estimator.

10.3.3 Properties

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

- IV estimators can have large standard errors. (small $R_{x,z}^2$)
- IV estimators can have large asymptotic bias if $Corr(z, u) \neq 0$ (we cannot check this formally).

For IV estimator,

$$plim \hat{\beta}_1 = \beta_1 + \frac{Corr(z, u)\sigma_u}{Corr(z, x)\sigma_x}$$

compared with OLS estimator

$$plim \tilde{\beta}_{OLS} = \beta_{OLS} + Corr(x, u)\frac{\sigma_u}{\sigma_x}$$

Remark 10.3. The R^2 in IV estimation can be negative, and we should be careful about interpreting R^2 in IV estimation.

10.4 IV in Multiple Regression

Consider the multiple regression model on k predictors, where y_2 is endogenous,

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

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 (10),

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

And solving above $k+1$ equations gives 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 $H_1 : \pi_k \neq 0$.

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