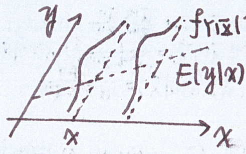
Classification: 1. Econometric Theory 2. Applied Econometric
(statistic x) nonexperimental economic data

statistic methods for estimating economic relationship, testing economic theories and evaluating and implementing government and business policy.

Axioms: 1. any economy ⇒ stochastic process, by probability law
2. economic phenomenon: stochastic DGP, realization

Types of Data: 1. Cross-Sectional Data (multi-entities, single period)
2. Time Series Data (single entity, multi-periods)
3. Panel Data (multi-entities, multi-periods) / longitudinal data.
4. Pooled cross section: pooling cross sections, diff years



- joint pdf: $f(x,y)$
- marginal pdf for X: $f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy$
- conditional pdf of Y given X: $f_{Y|X}(y|x)=\frac{f(x,y)}{f(x)}$
  - completely describe how Y depends on X.

$\begin{cases}\text{conditional mean}: E(Y|X)=\int_{-\infty}^{+\infty}y f_{Y|X}(y|x)dy\\ \text{conditional variance}: Var(Y|X)=E(Y^2|X)-[E(Y|X)]^2\end{cases}$

Causal Relationship: Ceteris Paribus (counterfactual reasoning)

Chapter 2 Simple Linear Regression Model

simple regression model: $y=\beta_0+\beta_1x+u$ (u: error term/disturbance)
population regression function (PRF): $E(y|x)=\beta_0+\beta_1x$ (Ass. $E(u|x)=0$)
observation: $y_i=\beta_0+\beta_1x_i+u_i$ (Ass. $E(u_i|x_i)=0$).

Ordinary Least Squares (OLS) ⇒ estimate β_0 and β_1.
Minimize objective function: $SSR(\beta_0,\beta_1)\equiv\sum_{i=1}^n(y_i-\beta_0-\beta_1x_i)^2$
FOC (first order conditions) / normal equations:

$$\begin{cases}\partial SSR(\hat\beta_0,\hat\beta_1)/\partial\beta_0=-2\sum_{i=1}^n(y_i-\hat\beta_0-\hat\beta_1x_i)=0 \Rightarrow \sum_{i=1}^n y_i=n\hat\beta_0+\hat\beta_1\sum_{i=1}^n x_i\\ \partial SSR(\hat\beta_0,\hat\beta_1)/\partial\beta_1=-2\sum_{i=1}^n(y_i-\hat\beta_0-\hat\beta_1x_i)x_i=0\Rightarrow\sum_{i=1}^n x_iy_i=\hat\beta_0\sum_{i=1}^n x_i+\hat\beta_1\sum_{i=1}^n x_i^2\end{cases}$$

plug in $\beta_0=\bar y-\hat\beta_1\bar x$ ⇒ unique solution to the convex minimization problem

$$\Rightarrow\begin{cases}\hat\beta_0=\bar y-\hat\beta_1\bar x\\ \hat\beta_1=\frac{\sum_{i=1}^n(x_i-\bar x)(y_i-\bar y)}{\sum_{i=1}^n(x_i-\bar x)^2}=\frac{\text{Sample }Cov(x,y)}{\text{Sample }Var(x)}\end{cases}\quad = \frac{\sum_{i=1}^n(x_i-\bar x)y_i}{\sum_{i=1}^n(x_i-\bar x)^2}=\frac{\sum_{i=1}^n x_iy_i-n\bar x\bar y}{\sum_{i=1}^n x_i-n\bar x^2}$$

condition (minimal): $S_x^2\equiv\sum_{i=1}^n(x_i-\bar x)^2>0$

Sample Regression Line / fitted line: $\hat y=\hat\beta_0+\hat\beta_1x$; Fitted Value $\hat y_i=\hat\beta_0+\hat\beta_1x_i$
Residual: $\hat u_i=y_i-\hat y_i=y_i-\hat\beta_0-\hat\beta_1x_i$ (estimate error term u_i; difference between · & l)

Properties (OLS): 1. $\bar{\hat u}_i\equiv\sum_{i=1}^n\hat u_i/n=0$ (mimic population moment condition $E(u)=0$). 1
2. $\bar y=\frac1n\sum_{i=1}^n y_i=\frac1n\sum_{i=1}^n\hat y_i+\frac1n\sum_{i=1}^n\hat u_i=\frac1n\sum_{i=1}^n\hat y_i=\bar{\hat y}$ — $E(u)=0$ but $\sum_{i=1}^n u_i=0$ ✗
3. $\sum_{i=1}^n x_i\hat u_i=0$. (FOC2), $\sum_{i=1}^n\hat y_i\hat u_i=\sum_{i=1}^n(\hat\beta_0+\hat\beta_1x_i)\hat u_i=0$ — (*) $(\hat\sigma)'=c\cdot\hat\sigma$, $R'=R$
4. pass through $(\bar x,\bar y)$. ★ Interpretation $(cy_i)=(c\beta_0)+(\frac{c\beta_1}{a})(ax_i)+(cu_i)$

Goodness of fit: $y_i=\hat y_i+\hat u_i$ — intercept included ⇒ $\sum_{i=1}^n\hat u_i=0\Rightarrow\bar{\hat y}=\bar y\Rightarrow SST=SSE+SSR$

$$\begin{cases}SST\equiv\sum_{i=1}^n(y_i-\bar y)^2\\ SSE\equiv\sum_{i=1}^n(\hat y_i-\bar{\hat y})^2\\ SSR\equiv\sum_{i=1}^n(y_i-\hat y_i)^2=\sum_{i=1}^n\hat u_i^2\end{cases}$$

$R_1^2=1-\frac{SSR}{SST}$, $R_2^2=\frac{SSE}{SST}$

(intercept ✓) $SST=SSE+SSR$, $0\le R_1^2=R_2^2\le1\Rightarrow\sum_{i=1}^n\hat u_i=0$, $\bar{\hat y}=\bar y$
$SST=\sum_{i=1}^n(y_i-\bar y)^2=\sum_{i=1}^n[y_i-\hat y_i+\hat y_i-\bar y]^2=SSR+\sum_{i=1}^n(\hat y_i-\bar{\hat y})^2+2\sum_{i=1}^n\hat u_i(\hat y_i-\bar y)$
(intercept ✗) $R_1^2\le1$, $R_2^2\ge0$

$\underline{R^2=[\widehat{Corr}(y,\hat y)]^2}$ $SSE=\sum_{i=1}^n(\hat y_i-\bar{\hat y})^2=\sum_{i=1}^n(\hat y_i-\bar{\hat y})(\hat y_i-\bar y)=(y_i-\hat u_i-\bar y)$
$=\sum_{i=1}^n(\hat y_i-\bar{\hat y})(y_i-\bar y)-\sum_{i=1}^n\hat u_i(\hat y_i-\bar{\hat y}_i)$

$R^2=\frac{SSE\cdot SSE}{SSE\cdot SST}=\frac{[\sum_{i=1}^n(\hat y_i-\bar{\hat y})(y_i-\bar y)]^2}{\sum_{i=1}^n(\hat y_i-\bar{\hat y})^2\sum_{i=1}^n(y_i-\bar y)^2}$ → sample ✓ but not 'real world'

Sampling Properties

(x_i, y_i) iid
- **SLR1** Linear in Params: $y=\beta_0+\beta_1x+u$ — $\sum_{i=1}^n k_i=0$, $\sum_{i=1}^n k_ix_i=0$
- **SLR2** Random Sampling: $y_i=\beta_0+\beta_1x_i+u_i$ — $E(y_i|\vec x)=E(y_i|x_i)$
- **SLR3** Sample Variation in Explantory variable
- **SLR4** Zero Conditional Mean: $E(u|x)=0$; $E(u_i|x_i)=0$

Law of Iterated Expectations (LIEs)
$\begin{cases}E[G(X,Y)]=E\{E[G(X,Y)|X]\}\\ E(Y)=E[E(Y|X)]\end{cases}$

$w_i=\frac{x_i-\bar x}{\sum_{j=1}^n(x_j-\bar x)^2}$, $[\sum_{i=1}^n w_i=0,\ \sum_{i=1}^n w_ix_i=1]$
$k_i=\frac1n-w_i\bar x$, $\sum_{i=1}^n w_i^2=\frac{1}{\sum_{j=1}^n(x_j-\bar x)^2}=\frac{1}{SST_x}$

1. unbiasedness. conditional unbiasedness $E(\hat\beta|\vec x)=\beta\Rightarrow$ unbiasedness $E(\hat\beta)=\beta$

$\hat\beta_1=\frac{\sum_{i=1}^n(x_i-\bar x)y_i}{\sum_{j=1}^n(x_j-\bar x)^2}=\sum_{i=1}^n\frac{x_i-\bar x}{\sum_{j=1}^n(x_j-\bar x)^2}y_i=\sum_{i=1}^n w_iy_i$. $E(\hat\beta_1|\vec x)=\sum_{i=1}^n E(w_iy_i|\vec x)=\sum_{i=1}^n w_iE(y_i|\vec x)\overset{SLR2\ iid}{=\!=}$
$\sum_{i=1}^n w_iE(y_i|x_i)=\sum_{i=1}^n w_i(\beta_0+\beta_1x_i)=\beta_0\sum_{i=1}^n w_i+\beta_1\sum_{i=1}^n w_ix_i=\beta_1$
$E(\hat\beta_0|\vec x)=E(\bar y-\hat\beta_1\bar x|\vec x)=E(\bar y|\vec x)-E(\hat\beta_1|\vec x)\cdot\bar x=E(\beta_0+\beta_1\bar x+\bar u|\vec x)-\beta_1\bar x=\beta_0$. $(\bar u=\frac1n\sum_{i=1}^n u_i)$

Direction of OVB: $\beta_2\cdot Corr(x_1,x_2)$

2. Variance. $\begin{cases}\text{best unbiased estimator / efficient estimator (smallest var)}\\ \text{best linear unbiased estimator (BLUE): proof under SLR1-5}\end{cases}$ — Gauss-Markov Theo.

- **SLR5** Conditional homoskedasticity: $Var(u|x)=\sigma^2$
  - SLR2+SLR5 ⇒ $Var(u_i|\vec x)=Var(u_i|x_i)=\sigma^2$
  - $E(y|x)=\beta_0+\beta_1x\Rightarrow Var(y|x)=\sigma^2$
  - op. conditional heteroskedasticity. $Var(y|x)$ depends on x.

$Var(u|x)=E(u^2|x)-E(u|x)=E(u^2)$
$=Var(u)=\sigma^2$ (error variance)
σ: standard deviation of the error

$\hat\beta_0=\bar y-\hat\beta_1\bar x=(\beta_0+\beta_1\bar x+\bar u)-\hat\beta_1\bar x=(\beta_0+\bar u)-(\hat\beta_1-\beta_1)\bar x$

$\hat\beta_1=\sum_{i=1}^n w_iy_i=\sum_{i=1}^n w_i(\beta_0+\beta_1x_i+u_i)=\beta_1+\sum_{i=1}^n w_iu_i$ ∴ $Var(\hat\beta_1|\vec x)=\sum_{i=1}^n Var(w_iu_i|\vec x)$
$=\sum_{i=1}^n w_i^2Var(u_i|x_i)=\sigma^2\sum_{i=1}^n w_i^2=\frac{\sigma^2}{SST_x}\rightarrow\boxed{Var(\hat\beta_1|\vec x)=\frac{\sigma^2}{SST_x}}$, $sd(\hat\beta_1)=\frac{\sigma}{\sqrt{SST_x}}$

$\frac{\sigma^2\sum_{i=1}^n x_i^2}{n\sum_{j=1}^n(x_j-\bar x)^2}$ — $\underline{Bias(\tilde\beta_1|\vec x)}=\hat\beta_2\tilde\delta_1$

$\hat\beta_0=\bar y-\hat\beta_1\bar x=\frac1n\sum_{i=1}^n y_i-\bar x\sum_{i=1}^n w_iy_i=\sum_{i=1}^n(\frac1n-\bar xw_i)y_i=\sum_{i=1}^n k_iy_i$. ∴ $\boxed{Var(\hat\beta_0|\vec x)}=\sum_{i=1}^n k_i^2Var(y_i|x_i)$
$=\sigma^2\sum_{i=1}^n k_i^2=\sigma^2\sum_{i=1}^n(\frac1n-w\bar x)^2=\sigma^2\sum_{i=1}^n(\frac{1}{n^2}+\bar x^2w_i^2-\frac{2\bar x}{n}w_i)=\sigma^2\sum_{i=1}^n(\frac{1}{n^2}+\bar x^2w_i^2)=\boxed{\sigma^2(\frac1n+\frac{\bar x^2}{SST_x})}=\frac{\sigma^2\sum_{i=1}^n x_i^2}{nSST_x}$

$\boxed{Cov(\hat\beta_0,\hat\beta_1|\vec x)}=Cov(\sum_{i=1}^n k_iy_i,\sum_{i=1}^n w_iy_i|\vec x)=Cov(\sum_{i=1}^n k_i(\beta_0+\beta_1x_i+u_i),\sum_{i=1}^n w_i(\beta_0+\beta_1x_i+u_i)|\vec x)$
$=Cov(\sum_{i=1}^n k_iu_i,\sum_{i=1}^n w_iu_i|\vec x)=\sum_i\sum_j Cov(k_iu_i,w_ju_j|\vec x)=\sum_i\sum_j k_iw_jCov(u_i,u_j|\vec x)$
$=\sum_i k_iw_iVar(u_i|\vec x)=\sigma^2\sum_i k_iw_i=\sigma^2\sum_i(\frac1n-w_i\bar x)w_i=-\sigma^2\bar x\sum_{i=1}^n w_i^2=\boxed{\frac{-\sigma^2\bar x}{SST_x}}$

Estimating σ^2: $\hat\sigma^2=\frac{1}{n-2}\sum_{i=1}^n\hat u_i^2=\frac{SSR}{n-2}$. $E(\hat\sigma^2|\vec x)=\sigma^2$ (conditionally unbiased)
$\hat\sigma=\sqrt{\hat\sigma^2}$: standard error (s.e.) of the regression / SER
Thus, $se(\hat\beta_1)=\frac{\hat\sigma}{\sqrt{SST_x}}$, $se(\hat\beta_0)=\hat\sigma\sqrt{\frac1n+\frac{\bar x^2}{SST_x}}$

*MSE = mean squared error

$\begin{cases}E(y-\beta_0-\beta_1x)=0\Rightarrow\beta_0=E(y)-\beta_1E(x)\\ E[x(y-\beta_0-\beta_1x)]=0\Rightarrow Cov(y-\beta_0-\beta_1x,x)=0\Rightarrow Cov(x,y)=\beta_1Var(x)\Rightarrow\beta_1=\frac{Cov(x,y)}{Var(x)}\end{cases}$ | $E(u_ix_i)=0$.

(moment conditions) / population moment restrictions
insights. true values of the params associate with moments of y and x, which are not estimates.

!! Equivalence of the MM estimators and OLS estimators !!
(replace the population moment restrictions by sample moment restrictions)

$\begin{cases}\frac1n\sum_{i=1}^n(y_i-\hat\beta_0-\hat\beta_1x_i)=0\\ \frac1n\sum_{i=1}^n x_i(y_i-\hat\beta_0-\hat\beta_1x_i)=0\end{cases}$

Sample analog principle: replace expectations by sample means

Regression Through the Origin: $y=\beta_1x+u$ — $\tilde\beta_1=\beta_1+\beta_0\frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2}$ conditionally biased
Minimize objective function: $SSR(\beta_1)\equiv\sum_{i=1}^n(y_i-\beta_1x_i)^2$
FOC: $\partial SSR(\tilde\beta)/\partial\beta_1=-2\sum_{i=1}^n x_i(y_i-\tilde\beta x_i)=0\Rightarrow\tilde\beta=\frac{\sum_{i=1}^n x_iy_i}{\sum_{i=1}^n x_i^2}$

define $R^2=1-\frac{\sum_{i=1}^n(y_i-\tilde\beta x_i)^2}{\sum_{i=1}^n y_i^2}$ (uncentered R^2) — properties: $\sum_{i=1}^n x_i\hat u_i=0$ (also $\sum_{i=1}^n\hat y_i\hat u_i=0$).
$\tilde R^2=1-\frac{\sum_{i=1}^n(y_i-\tilde\beta x_i)^2}{\sum_{i=1}^n(y_i-\bar y)^2}$ could be negative ($\bar y$ is a better fit)

Binary Variable (dummy variable)

$\begin{cases}E(y|x=0)=\beta_0\\ E(y|x=1)=\beta_0+\beta_1\end{cases}\Rightarrow\begin{cases}\hat\beta_0=\bar y_0: \text{average for non-}x\ (\bar y_0)\\ \hat\beta_1=\bar y_1-\bar y_0: \text{average for has-}x\ (\bar y_1).\end{cases}$

$y_i=\bar y_1x_i+\bar y_0(1-x_i)$. $x_i=1$ indicates receiving treatment
Average Treatment / Causal Effect (ATE, ACE) $\tau=E(\bar y_1)-E(\bar y_0)$
x independent of $\{\bar y_1,\bar y_0\}$: $E(y_i|x_i)=E[\bar y_0+(\bar y_1-\bar y_0)x_i|x_i]$
$=E[\bar y_0|x_i]+E[(\bar y_1-\bar y_0)|x_i]x_i=E(\bar y_0)+E(\bar y_1-\bar y_0)x_i\equiv\alpha_0+\tau x_\tau$
τ: ATE of interest. (treatment effect)

Chapter 3 Multiple Regression Analysis

Multiple regression model: $y=\beta_0+\beta_1x_1+\beta_2x_2+\cdots\beta_kx_k+u$
$SSR(\beta_0,\beta_1\cdots\beta_k)\equiv\sum_{i=1}^n(y_i-\beta_0-\beta_1x_{i1}-\beta_2x_{i2}-\cdots\beta_kx_{ik})^2$
FOCs properties: $\sum_{i=1}^n\hat u_i=0$, $\sum_{i=1}^n x_{ik}\hat u_i=0$ $(k=1,2,\cdots k)$, $\sum_{i=1}^n\hat u_i\hat y_i=0$, $\bar y=\bar{\hat y}$.
(MM interpretation): $E(u|x_1,x_2,\cdots x_k)=0\Rightarrow E(u)=0$, $E(u|x_k)=0$

Goodness of fit
adjusted-R^2: $\bar R^2=1-\frac{SSR/(n-k-1)}{SST/(n-1)}=1-\frac{n-1}{n-k-1}\frac{SSR}{SST}<R^2$ — It can be negative! (intercept included)

"Partialling Out" Interpretation

$$\begin{tikzcd} & x_1 \arrow[dl, "\beta_1"'] \arrow[d, "\delta_1"] \\ y & x_2 \arrow[l, "\beta_2"] \end{tikzcd}$$

regression model: $y_i=\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+u_i$
1. full regression, run y_i on $(1,x_{i1},x_{i2})\Rightarrow\hat\beta_0,\hat\beta_1,\hat\beta_2$
2. partialling out regression. $\begin{cases}\text{regress } x_{i1} \text{ on } (1,x_{i2})\Rightarrow\text{residuals } \hat r_{i1}\\ \text{regress } y_i \text{ on } \hat r_{i1}\Rightarrow\hat\beta_1\end{cases}$ $\sum_{i=1}^n\hat r_{i1}=\sum_{i=1}^n\hat r_{i1}x_{ik}=0$, $k\neq1$.

$\boxed{\hat\beta_1=\frac{\sum_{i=1}^n\hat r_{i1}y_i}{\sum_{i=1}^n\hat r_{i1}^2}}$ 1 and 2 are equivalent (Frisch-Waugh Theorem)

properties: $\hat\beta_1=\beta_1+\frac{\sum_{i=1}^n\hat r_{i1}u_i}{\sum_{i=1}^n\hat r_{i1}^2}$

1. SLR vs MLR

$\begin{cases}\tilde y_i=\tilde\beta_0+\tilde\beta_1x_{i1}\\ \hat y_i=\hat\beta_0+\hat\beta_1x_{i1}+\hat\beta_2x_{i2}\end{cases}$

$\boxed{\text{Thus}: \tilde\beta_1=\hat\beta_1+\hat\beta_2\tilde\delta_1}$
where $\tilde\delta_1$: reg x_{i2} on $(1,x_{i1})$:
$x_{i2}=\delta_0+\delta_1x_{i1}+v_i$

we have $\hat u_i=y_i-\hat\beta_0-\hat\beta_1x_{i1}-\hat\beta_2x_{i2}$
$\tilde\beta_1=\frac{\sum_{i=1}^n(x_{i1}-\bar x_1)y_i}{\sum_{j=1}^n(x_{j1}-\bar x_1)^2}=\frac{\sum_{i=1}^n(x_{i1}-\bar x)(\hat\beta_0+\hat\beta_1x_{i1}+\hat\beta_2x_{i2}+\hat u_i)}{\sum_{j=1}^n(x_{j1}-\bar x_1)^2}$
$=\frac{1}{SST_{x_1}}(\hat\beta_1\sum_{i=1}^n(x_{i1}-\bar x_1)x_{i1}+\hat\beta_2\sum_{i=1}^n(x_{i1}-\bar x_1)x_{i2})$
$=\hat\beta_1+\hat\beta_2\frac{\sum_{i=1}^n(x_{i1}-\bar x_1)x_{i2}}{SST_{x_1}}=\hat\beta_1+\hat\beta_2\tilde\delta_1$

2. Regression with k vs k-1 estimators

$\boxed{\tilde\beta_j=\hat\beta_j+\hat\beta_k\tilde\delta_j}$ $(j=1,2,\cdots k-1)$. where $\tilde\delta_j$: reg x_{ik} on $(1,x_{i1}\cdots x_{ik-1})$
$x_{ik}=\delta_0+\delta_1x_{i1}+\cdots\delta_{k-1}x_{ik-1}+v_i$ (error term)

proof: $\begin{cases}\tilde y_i=\tilde\beta_0+\tilde\beta_1x_{i1}+\cdots\tilde\beta_{k-1}x_{ik-1}\\ \hat y_i=\hat\beta_0+\hat\beta_1x_{i1}+\cdots\hat\beta_kx_{ik}\end{cases}$
by partialling-out regression: $\tilde\beta_1=\frac{\sum_{i=1}^n\hat r_{i1}y_i}{\sum_{i=1}^n\hat r_{i1}^2}$
where $\hat r_{i1}$: reg x_{i1} on $(1,x_{i2},\cdots x_{i,k-1})$: $\hat r_{i1}=x_{i1}-\hat\gamma_{10}-\hat\gamma_{12}x_{i2}-\cdots\hat\gamma_{1,k-1}x_{i,k-1}$
we have $\begin{cases}\sum_{i=1}^n\hat r_{i1}=\sum_{i=1}^n\hat r_{i1}x_{ij}=0,\ j=2,3,\cdots k-1\\ \sum_{i=1}^n\hat r_{i1}\hat u_i=\sum_{i=1}^n(x_{i1}-\hat\gamma_{10}-\hat\gamma_{12}x_{i2}-\cdots)\hat u_i=0\end{cases}$

$\tilde\beta_1=\frac{1}{\sum_{i=1}^n\hat r_{i1}^2}(\sum_{i=1}^n\hat r_{i1}(\hat\beta_0+\hat\beta_1x_{i1}+\cdots\hat\beta_kx_{ik}+u_i))=\hat\beta_1\frac{\sum_{i=1}^n\hat r_{i1}x_{i1}}{\sum_{i=1}^n\hat r_{i1}^2}+\hat\beta_k\frac{\sum_{i=1}^n\hat r_{i1}x_{ik}}{\sum_{i=1}^n\hat r_{i1}^2}$
$=\hat\beta_1(\frac{\sum_{i=1}^n\hat r_{i1}(x_{i1}-\hat\gamma_{10}-\hat\gamma_{12}x_{i2}-\cdots)}{\sum_{i=1}^n\hat r_{i1}^2})+\hat\beta_k\tilde\delta_1$ $(x_{ik}=\delta_0+\delta_1x_{i1}+\ \delta_{k-1}x_{ik-1}+v)$
$=\hat\beta_1+\hat\beta_k\tilde\delta_1$, same for $j=2,3,\cdots k-1$ — $\tilde\delta_1=\frac{\sum_{i=1}^n\hat r_{i1}x_{ik}}{\sum_{i=1}^n\hat r_{i1}^2}$ (partialling out)

Assumption for unbiasedness

iid
- **MLR1** Linear in Params: $y=\beta_0+\beta_1x_1+\beta_2x_2+\cdots u$
- **MLR2** Random Sampling: $y_i=\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\cdots u_i$
- **MLR3** No perfect Collinearity $<$ large F statistic for testing overall significance; many low t-statistic for testing individual slope coefficient
- **MLR4** Zero Conditional Mean: $E(u|x_1,x_2,\cdots x_n)=0$

3. The deletion and addition of few observations change the results significantly
4. possibly wrong sign

⇒ unbiasedness. conditional unbiasedness $E(\hat\beta_j|\{x_{i1},\cdots x_{ik}\}_{i=1}^n)=\beta_j$.

Inclusion of Irrelevant Variables (overspecify the model)
$E(\hat\beta_j)=\beta_j$ for $j=1,2,0$. $E(\hat\beta_3)=0\Rightarrow$ unbiased but variance ↑ ie $Var(\hat\beta_{1,2})$↑

Omitted Variables Bias (OVB) (underspecified the model) Dominate

$\begin{cases}\text{true model}: y_i=\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+u_i,\ E(u_i|x_{i1},x_{i2})=0\\ \text{SLR model}: y_i=\beta_0+\beta_1x_{i1}+v_i\ (v_i=\beta_2x_{i2}+u_i)\end{cases}$

$n\to+\infty$, $Var(\hat\beta)$, $Var(\tilde\beta)\to0$
(*) $MSE(\tilde\beta|\vec x)=[Bias(\tilde\beta|\vec x)]^2+Var(\tilde\beta|\vec x)$

$\tilde\beta_1=\frac{1}{SST_{x_1}}(\sum_{i=1}^n(x_{i1}-\bar x_1)(y_i-\bar y))=\frac{1}{SST_x}(\sum_{i=1}^n(x_{i1}-\bar x)(\beta_1(x_{i1}-\bar x)+(v_i-\bar v)))$
$=\beta_1+\frac{1}{SST_{x_1}}(\sum_{i=1}^n(x_{i1}-\bar x)v_i)=\beta_1+\beta_2\frac{\sum_i(x_{i1}-\bar x_1)x_{i2}}{\sum_i(x_{i1}-\bar x_1)^2}+\frac{\sum_i(x_{i1}-\bar x_1)u_i}{\sum_i(x_{i1}-\bar x_1)^2}$
$\boxed{E(\tilde\beta_1|\vec x)=\beta_1+\beta_2\tilde\delta_1}$ (Given $E(u_i|\vec x)=0$), $\tilde\delta_1$: reg x_{i2} on $(1,x_{i1})$

Variance (MLR1-5: Gauss-Markov assumptions)
- **MLR5** Conditional Homoskedasticity: $Var(u|x_1,\cdots x_k)=\sigma^2$
  - imply $Var(y|x_1,\cdots x_k)=\sigma^2$

$$\boxed{Var(\hat\beta_j|\vec x)=\frac{\sigma^2}{SST_j(1-R_j^2)},\ j=1,2,\cdots k,\ \text{where } SST_j=\sum_{i=1}^n(x_{ij}-\bar x_j)^2 \text{ and } R_j \text{ is the } R^2 \text{ from reg } x_j \text{ on other regressors.}\quad \vec x=\{x_{i1},\cdots x_{ik}\}_{i=1}^n}$$

$\hat\beta_1=\frac{\sum_{i=1}^n\hat r_{i1}y_i}{\sum_{i=1}^n\hat r_{i1}^2}$, $Var(\hat\beta_1|\vec x)=\frac{1}{\sum_{i=1}^n\hat r_{i1}^2}Var(\sum_{i=1}^n\hat r_{i1}y_i|\vec x)=\frac{1}{\sum_i\hat r_{i1}^2}\sum_{i=1}^n Var(\hat r_{i1}y_i|\vec x)$ MLR.2
$=\frac{1}{\sum_i\hat r_{i1}^2}\sum_{i=1}^n\hat r_{i1}Var(y_i|\vec x)=\frac{\sigma^2}{\sum_{i=1}^n\hat r_{i1}^2}=\frac{\sigma^2}{SST_1(1-R_1^2)}$. $(\sum_{i=1}^n\hat r_{i1}^2=SSR$ for x_{i1} on $(1,x_{i2}))$

- $R_j^2=1\Leftrightarrow$ perfect multicollinearity, variance → +∞

variance inflation factor (VIF): $VIF_j=\frac{1}{1-R_j^2}\Rightarrow Var(\hat\beta_j)=\frac{\sigma^2}{SST_j}VIF_j$
(*) high correlation between x_2,x_3 doesn't effect $Var(\hat\beta_1)$

Assumption MLR1-5 ⇒ OLS estimators are BLUE
(proof: 1 check linearity and unbiasedness, 2 minimum variance)

Estimating σ^2: $\hat{\sigma}$: standard error of the regression (SER)

$\sigma^2 = Var(u) = E(u^2)$, MLR.5, $\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots \hat{\beta}_k x_{ik}$

$\hat{\sigma}^2 = \frac{\sum_{i=1}^n \hat{u}_i^2}{n-k-1} = \frac{SSR}{n-k-1}$, denominator = degrees of freedom (df)

$\Rightarrow E(\hat{\sigma}^2|\vec{x}) = \sigma^2$, conditionally unbiased.

$se(\hat{\beta}_j) = \frac{\hat{\sigma}}{\sqrt{SST_j(1-R_j^2)}}$, standard error of $\hat{\beta}_j$ $\rightarrow sd(\hat{\beta}_j) = \sqrt{Var(\hat{\beta}_j|\vec{x})}$

(compared with $sd(\hat{\beta}_j) = \frac{\sigma}{\sqrt{SST_j(1-R_j^2)}}$, standard deviation of $\hat{\beta}_j$)

(*) $se(\hat{\beta}_j) = \frac{\hat{\sigma}}{\sqrt{n}\, sd(x_j)\sqrt{1-R_j^2}}$, where $sd(x_j) = \sqrt{\frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n}}$

■ Chapter 4 Multiple Regression Analysis: Finite Sample Inference

the T distribution: $T = \frac{Z}{\sqrt{X/n}}$ Z: standard normal distribution $\sim N(0,1)$
X: chi-square distribution with n df

Z and X independent ⇒ t distribution with n degrees of freedom

Let $X \sim t(n)$. $E(X) = 0$, if $n > 1$ $Var(X) = \frac{n}{n-2}$, if $n > 2$, $t(n) \rightarrow N(0,1), n \rightarrow +\infty$

the F distribution: $F = \frac{X_1/k_1}{X_2/k_2}$ X: chi-square distribution with k df
X_1, X_2 are independent

Simple Random Sampling: n objects selected at random from population

i.i.d. = independently and identically distributed 独立同分布

{ exact distribution / finit-sample distribution
approximate / asymptotic distribution (when $n \rightarrow +\infty$)

{ MLR6 (Normality): $u \sim N(0, \sigma^2)$ and is independent of $x_1, x_2, \cdots x_k$

MLR1-6: classical linear model (CLM) assumptions implies MLR5

⇒ OLS estimator is not only BLUE, but ~~the~~ minimum variance unbiased

(summarize): $y|\vec{x} \sim N(\beta_0 + \beta_1 x_1 + \cdots \beta_k x_k, \sigma^2)$ $\hat{\beta}_j = \beta_j + \sum_{i=1}^n w_{ij} u_i$, $w_{ij} = \frac{\hat{r}_{ij}}{SSR_j}$

$\Rightarrow \hat{\beta}_j|\vec{x} \sim N(\beta_j, Var(\hat{\beta}_j|\vec{x}))$ for $j = 0, 1, \cdots k$. $\Rightarrow \frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)}|\vec{x} \sim N(0,1)$ $sd(\hat{\beta}_j) = \sqrt{Var(\hat{\beta}_j|\vec{x})}$

---

the t Test Type I error will accumulate as we test more models ⇒ can't find true model

Terminology: { null hypothesis (H_0)
alternative hypothesis (H_1)
{ Type I error: reject H_0 when true (α)
Type II error: fail to reject H_0 when false (β)

size/significance level: probability of making type I error (α)

multicollinearity makes it less power ⇐ power: probability that a test correctly rejects the null ($1-\beta$)

test statistic (统计量): a function of the random sample

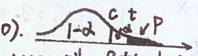
critical value: the value of test statistic when it just hit a given α.

rejection region: set of values of test statistic when rejects the null.
op. acceptance region / fail to reject region

the p-value: the smallest significance level to reject the null

Theorem: $\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)}|\vec{x} \sim t(n-k-1) = t_{df}$, called t statistic/ratio

1. One-sided Alternative ($H_0: \beta_j = 0$, $H_1: \beta_j > 0$).



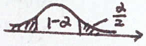
critical value $c_\alpha = t_\alpha(n-k-1)$: upper α-percentile of $t(n-k-1)$

rejection rule: rejection rule: $\underline{t_{\hat{\beta}_j} > c_\alpha}$

1. $|t_{\hat{\beta}_j}| > c_{\alpha/2}$ (*) interpretation: $c_{0.05,28} = 1.701 \Rightarrow P(t_{\hat{\beta}_j} > 1.701) = 0.05$

2. $P(|t_{n-k-1}| > t_{cal})$ p-value: reject when $\underline{p\text{-value} < \alpha}$ (eg. $P(t(28) > 2.763) = 0.005$)

3. CI 2 Two-sided Alternative ($H_0: \beta_j = 0$. $H_1: \beta_j \neq 0$)



(*) statistical significance: determined totally by t statistic
economic significance: focus on the magnitude of coefficients

Confidence Intervals (CI)

$-c_{\alpha/2} \leq \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \leq c_{\alpha/2} \Rightarrow \hat{\beta}_j - c_{\alpha/2} se(\hat{\beta}_j) \leq \beta_j \leq \hat{\beta}_j + c_{\alpha/2} se(\hat{\beta}_j)$ (two sided)

or: $\hat{\beta}_j \pm c_{\alpha/2} se(\hat{\beta}_j)$ in short

3 testing a single linear combination of the parameters

$H_0: \beta_1 = \beta_2$. $t = \frac{\hat{\beta}_1 - \hat{\beta}_2}{se(\hat{\beta}_1 - \hat{\beta}_2)}$, where $se(\hat{\beta}_1 - \hat{\beta}_2)$ is an estimato of $sd(\hat{\beta}_1 - \hat{\beta}_2)$

Type I error can accumulate when conducting more multiple tests.

$= \sqrt{Var(\hat{\beta}_1 - \hat{\beta}_2)} = \sqrt{Var(\hat{\beta}_1) + Var(\hat{\beta}_2) - 2Cov(\hat{\beta}_1, \hat{\beta}_2)} = \sqrt{se(\hat{\beta}_1)^2 + se(\hat{\beta}_2)^2 - 2S_{12}}$

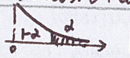
where S_{12} denotes an estimate of $Cov(\hat{\beta}_1, \hat{\beta}_2)$

⇓ def $\hat{\theta} = \hat{\beta}_1 - \hat{\beta}_2$, $CI = [\hat{\theta} - c_{\alpha/2} se(\hat{\theta}), \hat{\theta} + c_{\alpha/2} se(\hat{\theta})]$

---

the F Test

{ unrestricted regression
restricted regression (q restrictions).



⇓ $F = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)} \sim F(q, n-k-1)$ (MLR1-MLR6).

rejection rule: $F > F_\alpha(q, n-k-1)$ or p-value $\equiv P(F(q, n-k-1) > F) < \alpha$

Theorem: $F = \frac{(R_{ur}^2 - R_r^2)/q}{(1-R_{ur}^2)/(n-k-1)}$. q = number of restrictions under the null
n-k-1: df of the unrestricted model.

Testing the overall significance of MLR: $F = \frac{R^2/k}{(1-R^2)/(n-k-1)}$

- when $q=1$, $F(1, n-k-1) = [t(n-k-1)]^2$, or $F = t^2$
- when $q=2$, $F = \frac{1}{2} \frac{t_1^2 + t_2^2 - 2\hat{\rho}_{12} t_1 t_2}{1 - \hat{\rho}_{12}^2}$, where $\hat{\rho}_{12}$ is the estimator of the correlation between t_1 and t_2. when $\hat{\rho}_{12} = 0$, $F = \frac{1}{2}(t_1^2 + t_2^2)$

Remarks: 1. F-test can be used to test multiple linear restrictions of several regression coefficients. (actually T test does)
2. can be used to test whether one regression coefficient is equal to another. (actually T test does).

The law of iterated expectations.

1. $E(u_i|x_{i1}, x_{i2}) = 0 \Rightarrow E(u_i|x_{i1}) = 0$ (But not vice versa)
2. $E(u_i|x_{i1}, x_{i2}) = 0 \Rightarrow Cov(u_i, x_{i1}) = Cov(u_i, x_{i2}) = 0$ ($\forall m(x_{i1}), m(x_{i2})$ also holds)

Basic properties of χ^2 random variables: $X \sim \chi^2(m)$, $Y \sim \chi^2(n)$.

1. $E(X) = m$, $E(Y) = n$, $Var(X) = 2m$, $Var(Y) = 2n$. 2. $X + Y \sim \chi^2(m+n)$, X, Y independent

level-log: $y = \beta_0 + \beta_1 \log x + u$: $\Delta y = \beta_1 \cdot \frac{dx}{x} \Rightarrow \Delta y = \frac{\beta_1}{100}(\%\Delta x)$

log-level: $\log(y) = \beta_0 + \beta_1 x + u$: $\frac{1}{y}\Delta y = \beta_1 dx \Rightarrow (\%\Delta y) = 100\beta_1 \Delta x$.

log-log: $\log(y) = \beta_0 + \beta_1 \log x + u$: $\frac{1}{y}\Delta y = \beta_1 \cdot \frac{dx}{x} \Rightarrow (\%\Delta y) = \beta_1(\%\Delta x)$ elasticity.

- HW1-6 (Single Linear Regression: $y_i = \beta_0 + \beta_1 x_i + u_i$).

(1) proof: $Cov(\hat{\beta}_1, \bar{u}|\vec{x}) = 0$ (We have $\hat{\beta}_1 = \beta_1 + \sum_{i=1}^n w_i u_i$ already).

$Cov(\hat{\beta}_1, \bar{u}|\vec{x}) = Cov(\beta_1 + \sum_{i=1}^n w_i u_i, \frac{1}{n}\sum_{i=1}^n u_i|\vec{x}) = Cov(\sum_{i=1}^n w_i u_i, \frac{1}{n}\sum_{i=1}^n u_i|\vec{x})$

$= \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^n Cov(w_i u_i, u_j|\vec{x}) = \frac{1}{n}\sum_{i=1}^n Cov(w_i u_i, u_i|\vec{x}) = \frac{1}{n}\sum_{i=1}^n E(w_i u_i^2|\vec{x})$

$= \frac{1}{n}\sum_{i=1}^n w_i E(u_i^2|\vec{x}) = \frac{\sigma^2}{n}\sum_{i=1}^n w_i = 0$

(2) proof: $Var(\hat{\beta}_0|\vec{x}) = \sigma^2(\frac{1}{n} + \frac{\bar{x}^2}{SST_x})$. (We have $\hat{\beta}_0 = \beta_0 + \bar{u} - (\hat{\beta}_1 - \beta_1)\bar{x}$ already)

$Var(\hat{\beta}_0|\vec{x}) = Var(\bar{u} - (\hat{\beta}_1 - \beta_1)\bar{x}|\vec{x}) = Var(\bar{u}|\vec{x}) + Var(\bar{x}(\hat{\beta}_1 - \beta_1)|\vec{x}) - 2Cov(\bar{u}, (\hat{\beta}_1 - \beta_1)\bar{x}|$

$= \frac{1}{n^2}\sum_{i=1}^n Var(u_i|\vec{x}) + \bar{x}^2 Var(\hat{\beta}_1|\vec{x}) - 2Cov(\bar{u}, \hat{\beta}_1|\vec{x})\bar{x} = \frac{\sigma^2}{n^2} + \bar{x}^2 \cdot \frac{\sigma^2}{SST_x} \Rightarrow$ 得证.

(another proof is written on the other page).

- HW1-7 ($\tilde{\beta}_1 = \frac{\bar{y}^{(2)} - \bar{y}^{(1)}}{\bar{x}^{(2)} - \bar{x}^{(1)}}$. Q: calculate $Var(\tilde{\beta}_1|\vec{x})$). we have $\tilde{\beta}_1 = \beta_1 + \frac{\bar{u}^{(2)} - \bar{u}^{(1)}}{\bar{x}^{(2)} - \bar{x}^{(1)}}$ already).

proof: $Var(\tilde{\beta}_1|\vec{x}) = Var(\frac{\bar{u}^{(2)} - \bar{u}^{(1)}}{\bar{x}^{(2)} - \bar{x}^{(1)}}|\vec{x}) = \frac{1}{(\bar{x}^{(2)} - \bar{x}^{(1)})^2} Var(\bar{u}^{(2)} - \bar{u}^{(1)}|\vec{x}) = \frac{1}{(\bar{x}^{(2)} - \bar{x}^{(1)})^2}[Var(\bar{u}^{(2)}|\vec{x})$
$+ Var(\bar{u}^{(1)}|\vec{x}) - 2Cov(\bar{u}^{(2)}, \bar{u}^{(1)}|\vec{x})] = \frac{1}{(\bar{x}^{(2)} - \bar{x}^{(1)})^2} Var(\frac{2}{n}\sum_{i=1}^n u_i|\vec{x}) = \frac{4\sigma^2}{n(\bar{x}^{(2)} - \bar{x}^{(1)})^2}$

$\Rightarrow$ choosing $(\bar{x}^{(2)} - \bar{x}^{(1)})$ max.

- HW2-3 (multiple Linear Regression).

{ $y_i = \beta_0 + \beta_1 x_{i1} + u_i \Rightarrow \tilde{\beta}_1$
$y_i = \beta_0 + \beta_2 x_{i2} + u_i \Rightarrow \tilde{\beta}_2$
$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i \Rightarrow \hat{\beta}_1, \hat{\beta}_2$

we have $\hat{\beta}_1 = \frac{\sum_{i=1}^n \hat{r}_{i1} y_i}{\sum_{i=1}^n \hat{r}_{i1}^2}$, $\hat{r}_{i1}$: reg x_{i1} on $(1, x_{i2}) \Rightarrow$ residuals
$x_{i1} = \alpha_0 + \alpha_1 x_{i2} + r_{i1}$

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

when we have $Cov(x_1, x_2) = \sum_{i=1}^n (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2) = 0$. we have $\hat{r}_{i1} = x_{i1} - \hat{\alpha}_0 - \hat{\alpha}_1 x_1 = x_{i1} - \bar{x}_{i1}$

So we have $\hat{\beta}_1 = \tilde{\beta}_1$ under the condition where x_{i1} and x_{i2} $\underline{\text{aren't}}$ correlated.

- HW2-4 (MLR OVB).

{ $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i$
$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + v_i$

$\tilde{r}_i$: reg x_{i1} on $(1, x_{i2})$ residuals

$\tilde{\beta}_1 = \frac{\sum_{i=1}^n \tilde{r}_{i1} y_i}{\sum_{i=1}^n \tilde{r}_{i1}^2} = \frac{\sum_{i=1}^n \tilde{r}_{i1}(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i)}{\sum_{i=1}^n \tilde{r}_{i1}^2}$

$= \frac{1}{\sum_{i=1}^n \tilde{r}_{i1}^2}(\beta_1 \sum_{i=1}^n \tilde{r}_{i1} x_{i1} + \beta_3 \sum_{i=1}^n \tilde{r}_{i1} x_{i3} + \sum_{i=1}^n \tilde{r}_{i1} u_i)$

$= \beta_1 + \beta_3 \frac{\sum_{i=1}^n \tilde{r}_{i1} x_{i3}}{\sum_{i=1}^n \tilde{r}_{i1}^2} + \frac{\sum_{i=1}^n \tilde{r}_{i1} u_i}{\sum_{i=1}^n \tilde{r}_{i1}^2}$ (partialling-out).

$E(\tilde{\beta}_1|\vec{x}) = \beta_1 + \beta_3 \frac{\sum_{i=1}^n \tilde{r}_{i1} x_{i3}}{\sum_{i=1}^n \tilde{r}_{i1}^2} \neq \beta_1$. (*) notice the difference with deduction from page 1.

- HW2-5 (overall significance F-test).

Model 1: $y_i = \beta_0 + u_i$.

OLS estimator: FOC $-2\sum_{i=1}^n (y_i - \tilde{\beta}_0) = 0 \Rightarrow \tilde{\beta}_0 = \bar{y}$.

$\therefore SSR = \sum_{i=1}^n u_i^2 = \sum_{i=1}^n (y_i - \bar{y})^2 = SST$. $R^2 = 1 - \frac{SSR}{SST} = 0$.

Model 2: $y_i = \beta_0 + \beta_1 x_{i1} + \cdots \beta_k x_{ik} + u_i$

$F = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)} = \frac{[SST - SST(1-R^2)]/k}{SST(1-R^2)/(n-k-1)} = \frac{R^2/k}{(1-R^2)/(n-k-1)}$ (notice: $SSR = SST(1-R^2)$).

- Partialling-out Equation (MLR): $\hat{\beta}_1 = \frac{\sum_{i=1}^n \hat{r}_{i1} y_i}{\sum_{i=1}^n \hat{r}_{i1}^2}$

reg x_1 on $(1, x_2, \cdots x_k)$, $x_{i1} = \hat{x}_{i1} + \hat{r}_{i1}$

FOC: $\sum_{i=1}^n (\hat{x}_{i1} + \hat{r}_{i1})(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots \hat{\beta}_k x_{ik}) = 0$. (notice we have $\sum_{i=1}^n \hat{x}_{i1} \hat{u}_i = 0$) Because $\hat{x}_{i1}$ is a linear function of $x_{i2}, \cdots x_{ik}$.

$\Rightarrow \sum_{i=1}^n \hat{r}_{i1}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots \hat{\beta}_k x_{ik}) = 0$. (notice we have $\sum_{i=1}^n x_{ij} \hat{r}_{i1} = 0$) $j \neq 1$.

$\Rightarrow \sum_{i=1}^n \hat{r}_{i1}(y_i - \hat{\beta}_1 x_{i1}) = 0$ (notice we have $\sum_{i=1}^n \hat{x}_{i1} \hat{r}_{i1} = 0$) $\Rightarrow \sum_{i=1}^n \hat{r}_{i1}(y_i - \hat{\beta}_1 \hat{r}_{i1}) = 0 \Rightarrow \hat{\beta}_1 = \frac{\sum_{i=1}^n \hat{r}_{i1} y_i}{\sum_{i=1}^n \hat{r}_{i1}^2}$

- MLR Estimator - Unbiasedness

$\hat{\beta}_1 = \beta_1 + (\sum_{i=1}^n \hat{r}_{i1} u_i)/(\sum_{i=1}^n \hat{r}_{i1}^2)$. $E(\hat{\beta}_1|\vec{x}) = \beta_1 + E(\sum_{i=1}^n \hat{r}_{i1} u_i)/\sum_{i=1}^n \hat{r}_{i1}^2|\vec{x}) = \beta_1 + \frac{\sum_{i=1}^n \hat{r}_{i1} E(u_i|\vec{x})}{\sum_{i=1}^n \hat{r}_{i1}^2} = \beta_1$

---

critical values of the t distribution

				α		
1-tailed:		0.1	0.05	0.025	0.01	0.005
2-tailed:		0.2	0.1	0.05	0.02	0.1
	1	3.078	6.314	12.706	31.821	63.657
	5					
	10					
df	15					
	20					
	25					
	30	1.310	1.697	2.042	2.457	2.750
	60	1.296	1.671	2.000	2.390	2.660
	90	1.291	1.662	1.987	2.368	2.632
	120	1.289	1.658	1.980	2.358	2.617
	∞	1.282	1.645	1.960	2.326	2.576

1 Sample Distribution (estimator) and Consistency

{ exact approach (finit/small sample approach)
asymptotic approach (large sample approach)

Definition: (1) Consistency ($\text{plim}\,\hat\beta=\beta$) ⇒ LLN ⇔ unbiasedness (finit sample)

(2) Asymptotic Efficiency ($\text{asy.Var}(\hat\beta)=\min$) ⇒ CLT ⇔ BLUE (finit sample)

(1) consistency of the OLS estimators (MLR1-MLR4*, $E(u)=0$ & $Cov(x_j,u)=0$)

$$\hat\beta_1=\frac{\sum_{i=1}^n(x_{i1}-\bar x_1)y_i}{\sum_{i=1}^n(x_{i1}-\bar x_1)^2}=\beta_1+\frac{\frac1n\sum_{i=1}^n(x_{i1}-\bar x_1)u_i}{\frac1n\sum_{i=1}^n(x_{i1}-\bar x_1)^2}\xrightarrow{P}\beta_1+\frac{Cov(x_{i1},u_i)}{Var(x_{i1})}=\beta_1$$

$E(\hat\beta_1)|_{n\to+\infty}=\beta$ asymptotic unbiasedness

$$\hat\beta_0=\bar y-\hat\beta_1\bar x\xrightarrow{P}E(y_i)-\beta_1E(x_i)=E(y_i-\beta_1x_i)=E(\beta_0+u_i)=\beta_0$$

($E(u)=0$ ⇒ guarantee the usual marginal effect interpretation of β_j)

(2) inconsistency caused by OVB (*ALL OLS estimators are generally biased)

True model: $y_i=\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+u_i$, $E(u_i|x_{i1},x_{i2})=0$ ↳ $Cov(x_1,x_2)\neq0$

Run SLR: $y_i=\beta_0+\beta_1x_{i1}+v_i$. ($v_i=\beta_2x_{i2}+u_i$) ⇒ x_1 is correlated with v

$$\Rightarrow\hat\beta_1\xrightarrow{P}\beta_1+\frac{Cov(x_{i1},v_i)}{Var(x_{i1})}=\beta_1+\beta_2\frac{Cov(x_{i1},x_{i2})}{Var(x_{i1})}$$ (asymptotic bias): positive/negative

2 Asymptotic Normality and Large Sample Inference (MLR1-MLR5).

Recall MLR6 (Normality): $u\sim N(0,\sigma^2)$ and is independent of $x_1,x_2,\dots x_k$

⇒ for making exact inference when using t and F statistics

CLT Review: $Z_n\equiv\sqrt n(\bar X_n-\mu)\xrightarrow{d}N(0,\sigma^2)$ or $\frac{\bar X_n-\mu}{\sigma/\sqrt n}\xrightarrow{d}N(0,1)$

Slutsky Theorem: $Z_n\xrightarrow{d}Z$, $a_n\xrightarrow{P}a$, $b_n\xrightarrow{P}b$ ⇒ $a_n+b_nZ_n\xrightarrow{d}a+bZ$

CMT (central mapping theorem): $X_n\xrightarrow{d}X$, $g:R^k\to R$, continuous ⇒ $g(X_n)\xrightarrow{d}g(X)$

(1) $\sqrt n(\hat\beta_j-\beta_j)\xrightarrow{d}N(0,\frac{\sigma^2}{a_j^2})$, $\text{asy.Var}(\sqrt n(\hat\beta_j-\beta_j)):=\frac{\sigma^2}{a_j^2}=\lim_{n\to\infty}Var(\sqrt n(\hat\beta_j-\beta_j))$.

$a_j^2=\text{plim}\frac1n\sum_{i=1}^n\hat r_{ij}^2$. $\hat r_{ij}$ are residuals regressing x_j on other regressors.

(2) $\hat\sigma^2\xrightarrow{P}\sigma^2=Var(u)$ (proof: LLN). ($\hat\sigma^2=\frac{1}{n-k-1}\sum_{i=1}^n\hat u_i^2$) large sample "$\frac1n$" is OK.

(3) $\frac{\hat\beta_j-\beta_j}{sd(\hat\beta_j)}\xrightarrow{d}N(0,1)$ and $\frac{\hat\beta_j-\beta_j}{se(\hat\beta_j)}\xrightarrow{d}N(0,1)$. Comparasion (MLR.6): $(\hat\beta_j-\beta)/se(\hat\beta_j)\sim t(n-k-1)$

$Var(\hat\beta_j|\vec x)=\frac{\sigma^2}{\sum_i\hat r_{ij}^2}$

reg (x_j) on ($1,x_i$)

⇒ R_j, $SSR=\sum_{i=1}^n\hat r_{ij}^2$

$SSR_j=SST_j(1-R_j^2)$

$Var(\hat\beta_j|\vec x)=\frac{\sigma^2}{SST_j(1-R_j^2)}$ (MLR1-MLR5), $\frac1nSST_j=\frac1n\sum_{i=1}^n(x_{ij}-\bar x_j)^2\xrightarrow{P}Var(x_{ij})\equiv\sigma_j^2$

{ $sd(\hat\beta_j)=\sigma/\sqrt{SST_j(1-R_j^2)}$
$se(\hat\beta_j)=\hat\sigma/\sqrt{SST_j(1-R_j^2)}\xrightarrow{P}sd(\hat\beta_j)$

ALSO we have $\text{plim}(R_j)=\rho_j^2\in[0,1]$ (under MLR1-MLR4)

$$\Rightarrow se(\hat\beta_j)=\frac{\hat\sigma/\sqrt n}{\sqrt{\frac1nSST_j(1-R_j)^2}}\to\frac{\sigma/\sqrt n}{\sigma_j\sqrt{1-\rho_j^2}}=\frac{c_j}{\sqrt n},\text{ where }c_j=\frac{\sigma}{\sigma_j\sqrt{1-\rho_j^2}}$$

when SLR: $se(\hat\beta_1)=\frac{\hat\sigma/\sqrt n}{\sqrt{\frac1nSST_x}}\to\frac{\sigma}{\sigma_x}\cdot\frac{1}{\sqrt n}$

proof for (1): $S_x^2=\frac1n\sum_{i=1}^n(x_i-\bar x)^2$. $\hat\beta_1=\beta_1+\frac{\sum(x_i-\bar x)u_i}{\sum(x_i-\bar x)^2}$

$\Rightarrow\sqrt n(\hat\beta_1-\beta_1)=\frac{\sum_{i=1}^n(x_i-\bar x)u_i}{\sqrt nS_x^2}$, where we have $S_x^2\xrightarrow{P}Var(x_i)\equiv\sigma_x^2$ (LLN)

SLR: $\sqrt n(\hat\beta_1-\beta_1)\xrightarrow{d}N(0,\frac{\sigma^2}{\sigma_x^2})$

$$\frac{\sum(x_i-\bar x)u_i}{\sqrt n}=\frac{1}{\sqrt n}\sum_{i=1}^n\{[x_i-E(x_i)]+[E(x_i)-\bar x]\}u_i=\frac{1}{\sqrt n}\sum_{i=1}^n[x_i-E(x_i)]u_i+\frac{E(x_i)-\bar x}{\sqrt n}\sum_{i=1}^nu_i$$

$$A_n=\frac{1}{\sqrt n}\sum_{i=1}^n[x_i-E(x_i)]u_i\xrightarrow{d}N(0,Var[(x_i-E(x_i))u_i])=N(0,\sigma^2\sigma_x^2)$$

$$(Var([x_i-E(x_i)]u_i)=E([x_i-E(x_i)]^2u_i^2)-E^2([x_i-E(x_i)]u_i)$$
$$=E([x_i-E(x_i)]^2E(u_i^2|x_i))-0^2=\sigma^2E([x_i-E(x_i)]^2)=\sigma^2\sigma_x^2)$$

$$B_n=\frac{1}{\sqrt n}(E(x_i)-\bar x)\sum_{i=1}^nu_i\xrightarrow{d}0\cdot N(0,\sigma^2)=0.\text{ which implies }B_n\xrightarrow{P}0.$$

Thus $\frac{1}{\sqrt n}\sum_{i=1}^n(x_i-\bar x)u_i\xrightarrow{d}N(0,\sigma^2\sigma_x^2)$, $\sqrt n(\hat\beta_1-\beta_1)\xrightarrow{d}N(0,\sigma^2/\sigma_x^2)$

That is $\lim Var(\sqrt n(\hat\beta_1-\beta_1))=\sigma^2/\sigma_x^2$. (under SLR when $k=1$).

{ asymptotic t test $\sim N(0,1)$. under the null.
asymptotic F test $\sim F(q,\infty)$. under the null

Remember an alternative: Lagrange Multiplier (LM)

The LM Statistic: $y_i=\beta_0+\beta_1x_{i1}+\dots\beta_kx_{ik}+u_i$. $H_0:\beta_{k-q+1}=\dots\beta_k=0$ ⇒ reg $\tilde u_i$ on $x_{ik-q+1}\cdots x_k$ $R^2\approx0$. But we should run $\tilde u_i$ on all x_{ij})

restricted: $y_i=\tilde\beta_0+\tilde\beta_1x_{i1}+\dots\tilde\beta_{k-q}x_{ik-q}+\tilde u_i$. → obtain $\tilde u_i$

auxiliary regression: $\tilde u_i$ on $(1,x_1,\dots x_k)$ → obtain R_u^2 (correlated $x_i x_j$)

LM statistic: $LM=nR_u^2$, $LM\xrightarrow{d}\chi^2(q)$ under the null

reject the null: $LM>\chi_\alpha^2(q)$, or the p value $<\alpha$.

(*) Ideally, $\tilde u_i$ should not be correlated with $x_{i,k-q+1},\dots x_k$

3 Asymptotic Efficiency of OLS

(SLR5. conditional homoskedasticity: $Var(u|x)=\sigma^2$).

OLS is asymptotic efficient under the Gauss-Markov Assumptions (e.g. SLR)

$y=\beta_0+\beta_1x+u$, $E(u|x)=0$. also we have u is uncorrelated with $g(x)$.

Let $z_i=g(x_i)$, moment conditions:

{ $E(u_i)=E(y_i-\beta_0-\beta_1x_i)=0$
$E(u_iz_i)=E((y_i-\beta_0-\beta_1x_i)z_i)=0$

$\Rightarrow\beta_1=\frac{Cov(z_i,y_i)}{Cov(z_i,x_i)}$, $\tilde\beta_1=\frac{\sum_{i=1}^n(z_i-\bar z)y_i}{\sum_{i=1}^n(z_i-\bar z)x_i}$ ⇒ satisfy conditions.

where $\tilde\beta_1$ makes a consistent estimator of β_1 and is linear in y_i.

MLR: $\hat\beta_1=\frac{\sum_{i=1}^n\hat r_{i1}y_i}{\sum_{i=1}^n\hat r_{i1}^2}$ → $=\beta_1+\frac{\sum_{i=1}^n\hat r_{i1}u_i}{\sum_{i=1}^n\hat r_{i1}^2}$

reg x_{i1} on $(1,x_{i2})$ ⇒ $\hat r_{i1}$

$$Var(\tilde\beta_1|\vec x)=Var\left(\frac{\sum_{i=1}^n(z_i-\bar z)u_i}{\sum_{i=1}^n(z_i-\bar z)x_i}\Big|\vec x\right)=\frac{1}{(\sum(z_i-\bar z)x_i)^2}Var\left(\sum_{i=1}^n(z_i-\bar z)u_i|\vec x\right)$$

$$=\frac{1}{(\sum(z_i-\bar z)x_i)^2}\sum_{i=1}^n(z_i-\bar z)^2E(u_i^2|\vec x)=\frac{\sigma^2}{n}\cdot\frac{\frac1n\sum_{i=1}^n(z_i-\bar z)^2}{(\frac1n\sum_{i=1}^n(z_i-\bar z)x_i)^2}\approx\frac{\sigma^2}{n}\frac{Var(z_i)}{Cov(z_i,x_i)^2}$$

Thus $\text{Asy.Var}(\tilde\beta_1)=\frac{\sigma^2}{n}\cdot\frac{Var(z_i)}{Cov(z_i,x_i)^2}$, notice that $[Cov(z_i,x_i)]^2\leq Var(z_i)Var(x_i)$

For OLS estimator: $Var(\hat\beta_1|\vec x)=\frac{\sigma^2}{n}\cdot\frac{\frac1n\sum(x_i-\bar x)^2}{(\frac1n\sum(x_i-\bar x)x_i)^2}\approx\frac{\sigma^2}{n}\cdot\frac{1}{Var(x_i)}\equiv\text{Asy.Var}(\hat\beta_1)$

⇒ $\text{Asy.Var}(\hat\beta_1)\leq\text{Asy.Var}(\tilde\beta_1)$. for any choice of $g(x)=z$.

Chapter 6 Multiple Regression Analysis: Further Issues

1 Effects of Data Scaling and Beta Coefficients

change scale of y: coefficient, $se(\hat\beta)$, CI changes, R^2, p-values doesn't change

change scale of x: coefficient $\hat\beta_j$, $se(\hat\beta_j)$, CI changes, doesn't change others

standardization: $\frac{y_i-\bar y}{\hat\sigma_y}=\frac{\hat\sigma_1}{\hat\sigma_y}\hat\beta_1\cdot\frac{x_{i1}-\bar x_1}{\hat\sigma_1}+\dots\frac{\hat\sigma_k}{\hat\sigma_y}\hat\beta_k\cdot\frac{x_{ik}-\bar x_k}{\hat\sigma_k}+\frac{\hat u_i}{\hat\sigma_y}$

$y_i=\hat\beta_0+\hat\beta_1x_{i1}+\dots\hat\beta_kx_{ik}+\hat u_i$
$\bar y=\hat\beta_0+\hat\beta_1\bar x_{i1}+\dots\hat\beta_k\bar x_k$

let $\hat b_j=\frac{\hat\sigma_j}{\hat\sigma_y}\hat\beta_j$ for $1\dots k$. Z-scores $z_y=\hat b_1z_1+\hat b_2z_2+\dots\hat b_kz_k+\text{error}$

↑ standardized coefficient / beta coefficient

(1) reflect the standard deviation change of y for one change in x.

(2) scale irrelevant, put regressors on an equal footing

(3) no change in significance for the estimation on R^2

2 Functional Form (natural log, quadratic/cubic forms, interactions)

(1) log model: $\ln y=\beta_0+\beta_1\ln x+u$: elasticity (1% x → β_1% y)

$y=\beta_0+\beta_1\ln x+u$: 1% → $\frac{\beta_1}{100}$ y. $\ln y=\beta_0+\beta_1x+u$: 1 x → $(100\beta_1)$% y.

Exact percentage change: $\ln y=\beta_0+\beta_1\ln x_1+\beta_2x_2+u$. | $100\beta_1$: semi-elasticity of y respect to x.

$\frac{\Delta y}{y}=\exp(\beta_2\Delta x_2)-1$

$\%\Delta y=100[\exp(\beta_2\Delta x_2)-1]$ (fix x_1).

Remarks: 1. invariant of scale 2. give direct estimate of elasticity/semi-ε.
3. reduce heteroskedasticity and skewness. 4. limit outlier effect.
5. cannot be used when variable take 0 or negative values.
6. not legitimate to compare R^2 from y-model and log(y)-model

(2) Quadratics: $y=\beta_0+\beta_1x+\beta_2x^2+u$. estimated marginal effect $\Delta\hat y/\Delta x=\hat\beta_1+2\hat\beta_2x$

Remarks: 1. turning point (few observation ⇒ ignore that side)
2. reciprocal form: $y=\beta_0+\beta_1\frac1x+u$.

(3) Interaction Form: $y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_1x_2+u$. ⇒ model changing marginal effect

reparameterized: $y=\beta_0+\delta_1x_1+\delta_2x_2+\beta_3(x_1-\mu_1)(x_2-\mu_2)+u$ (δ_2: when $x_1=\mu_1$, the partial effect of x_2 on y. $\delta_2=\beta_2+\beta_3\mu_1$)

μ_1,μ_2 are the population mean of x_1 and x_2

3 More on R^2 and Selection of Regressors

$R^2=[\widehat{Corr}(y,\hat y)]^2$

R^2 statistic valid: 1. OLS estimators. 2. linear relationship estimated 3. intercept

{ $SST=\sum_{i=1}^n(y_i-\bar y)^2$
$SSE=\sum_{i=1}^n(\hat y_i-\bar{\hat y})^2$
$SSR=\sum_{i=1}^n\hat u_i^2$

$R^2=1-\frac{SSR}{SST}=1-\frac{SSR/n}{SST/n}$, adjusted $R^2:=\bar R^2=1-\frac{SSR/(n-k-1)}{SST/(n-1)}=1-\frac{\hat\sigma^2}{SST/(n-1)}$

population R-squared $\rho^2=1-\frac{\sigma_u^2}{\sigma_y^2}$. R^2 and $\bar R^2$ aren't unbiased estimator of ρ.

$\Rightarrow\bar R^2=1-(1-R^2)\frac{n-1}{n-k-1}$ (definition) $\hat\sigma^2=\frac{1}{n-k-1}\sum_{i=1}^n\hat u_i^2=\frac{1}{n-k-1}SSR$

(1) $\bar R^2$ chooses between nonnested models with same dependent variables. (diff. SST)

(2) F-test chooses between nested models to limit redundant regressors

Recall: $F=\frac{R^2/k}{(1-R^2)/(n-k-1)}$ (overall significance), $F=\frac{(R_{ur}^2-R_r^2)/q}{(1-R_{ur}^2)/(n-k-1)}=\frac{(SSR_r-SSR_{ur})/q}{SSR_{ur}/(n-k-1)}$

(3) adding regressors: multicollinearity, reduce error variance and $se(\hat\beta)$.

4 Prediction and Residual Analysis

measure uncertainty in predicted value: $\hat\theta_0=\hat y^c=\hat\beta_0+\hat\beta_1c_1+\dots\hat\beta_kc_k$.

rewrite $\beta_0=\theta_0-\beta_1c_1-\dots\beta_kc_k$. ⇒ $y=\theta_0+\beta_1(x_1-c_1)+\beta_2(x_2-c_2)+\dots\beta_k(x_k-c_k)+u$

regress y_i on $(1,x_{i1}-c_1,x_{i2}-c_2,\dots x_{ik}-c_k)$ ⇒ obtain $\hat\theta_0$ and $se(\hat\theta_0)$.

$E(y^0|x_1^0,x_2^0,\dots x_k^0)$.

CI: $(\hat\theta_0-c_{\alpha/2}\cdot se(\hat\theta_0),\hat\theta_0+c_{\alpha/2}\cdot se(\hat\theta_0))$. (c: t-test for small sample, normal for large)

↑ CI for prediction
↓ CI for outcome y^0

(*) when $c_j=\bar x_j$, $se(\hat\theta_0)$ will be the smallest. this method rely on the average sub-population

$y^0=\beta_0+\beta_1x_1^0+\dots\beta_kx_k^0+u^0$ (u^0 is the unobserved error)

OLS prediction: $\hat y^0=\hat\beta_0+\hat\beta_1x_1^0+\dots\hat\beta_kx_k^0$

prediction error: $\hat e^0=y^0-\hat y^0=(\beta_0+\beta_1x_1^0+\dots\beta_kx_k^0)+u^0-\hat y^0$

- SLR properties recalled:
  1. $\bar{\hat u}_i=0$, $\frac1n\sum_{i=1}^n\hat u_i=0$
  2. $\bar y=\frac1n\sum_{i=1}^ny_i=\frac1n\sum_{i=1}^n(\hat y_i+\hat u_i)=\bar{\hat y}$
  3. FOC2: $\sum_{i=1}^nx_i\hat u_i=0$ (normal equation)
- MLR properties
  1. $\sum_{i=1}^n\hat r_{i1}x_{i1}=\sum_{i=1}^n\hat r_{i1}^2$ $(=\hat r_{i1}(x_{i1}-\hat\alpha_0-\hat\alpha_1x_{i2}))$ $x_{i1}=\hat\alpha_0+\hat\alpha_1x_{i2}+\hat r_{i1}$

we have: $E(\hat y^0|\vec x,x^0)=E(\hat\beta_0+\hat\beta_1x_1^0+\dots\hat\beta_kx_k^0|\vec x,x^0)=\beta_0+\beta_1x_1^0+\dots\beta_kx_k^0$

$E(\hat e^0|\vec x,x^0)=E(u^0|\vec x,x^0)=0$ (random sampling).

$Var(\hat e^0|\vec x,x^0)=Var(\hat y^0|\vec x,x^0)+Var(u^0|\vec x,x^0)=Var(\hat y^0|\vec x,x^0)+\sigma^2$ $(Cov(u^0,\hat y^0)=0)$

estimate: $se(\hat e^0)=\sqrt{\hat\sigma^2+[se(\hat y^0)]^2}$ corresponds to $Var(\hat\beta|\vec x)$.

$\Rightarrow\frac{\hat e^0}{se(\hat e^0)}\sim t(n-k-1)$. CI: $(\hat y^0-t_{\alpha/2}(n-k-1)se(\hat e^0),\hat y^0+t_{\alpha/2}(n-k-1)se(\hat e^0))$.

(*) much wider than above method. $\hat\sigma\xrightarrow{P}\sigma^2$, $se(\hat y^0)\to\frac{c}{\sqrt n}$ (some constant c).

this means $\hat\sigma^2\gg[se(\hat y^0)]^2$ in the definition of $se(\hat e^0)$.

Residual Analysis. eg1 reg price on car characters: big negative residual ⇒ good deal
eg2 reg earnings on student characters: big positive residual ⇒ value-added

predict y in log model: $\log(y)=\beta_0+\beta_1x_1+\dots\beta_kx_k+u$

(1) $\widehat{\log(y)}=\hat\beta_0+\hat\beta_1x_1+\dots\hat\beta_kx_k$, i.e. $\hat y=\exp(\widehat{\log(y)})$ ⇒ systematically underestimate $E(\hat y)$.

Jensen Inequality: $E[\varphi(Y)]\geq\varphi[E(Y)]$ for any convex function $\varphi(\cdot)$

Thus $E[\log(Y)|x_1,\dots x_k]=\beta_0+\beta_1x_1+\dots\beta_kx_k$

$E[y|x_1\dots x_k]\geq\exp(E[\log(Y)|x_1\dots x_k])=\exp(\beta_0+\beta_1x_1+\dots\beta_kx_k)$

(2) $E(y|\vec x)=E(\exp(\beta_0+\beta_1x_1+\dots\beta_kx_k)\exp(u)|\vec x)=\exp(\beta_0+\beta_1x_1+\dots\beta_kx_k)\cdot E(\exp(u))\equiv m\cdot\alpha_0$.

m can be estimated by $\widehat{\log(y)}$: $m\sim\exp(\widehat{\log y})$. $\alpha_0=E(e^u)$

α_0: $E(\exp(u))>\exp(E(u))=\exp(0)=1$.

biased, consistent 1. $u\sim N(0,\sigma^2)$. $\alpha_0=\exp(\frac{\sigma^2}{2})$. Thus $\alpha_0\sim\hat\alpha_0=\exp(\frac{\hat\sigma^2}{2})\Rightarrow\hat y=e^{\frac{\hat\sigma^2}{2}}e^{\widehat{\log y}}$

biased, consistent 2. sample analogue $\hat\alpha_0=\frac1n\sum_{i=1}^n\exp(\hat u_i)$ ⇒ consistent

biased, consistent 3. regress y_i on $\hat m_i=\exp(\hat\beta_0+\hat\beta_1x_{i1}+\dots\hat\beta_kx_{ik})$. $E(y|\vec x)=m\cdot\alpha_0\Rightarrow\hat\alpha_0=\frac{\sum_{i=1}^n\hat m_iy_i}{\sum_{i=1}^n\hat m_i^2}$.

(*) compare correlation (fitted value from auxiliary reg, y) and R^2 from level regression to choose from level or log method.

Chapter 7 Multiple Regression Analysis: Qualitative Information

$y=\beta_0+\beta_1x+\delta_0d+u$, $d=0/1$. ($d=0$: base/benchmark group). $\delta_0=E(y|d=1,x)-E(y|d=0,x)$

$\log(y)=\beta_0+\beta_1d+u$. $\Rightarrow\beta_1=\frac{\partial\log(y)}{\partial d}\approx\frac{\Delta y}{y}$, or more precisely: $\Delta y\%=100[\exp(\beta_1)-1]$

1. Dummy Variables for Multiple Categories

g groups: at most g-1 dummy variables (intercept ~ base group) ⇒ avoid perfect collinear

[EG]: $Wage=\beta_0+\delta_0female+\delta_1married+\beta_1educ+\dots+u$. ⇒ marriage premium same for genders

⇒ $Wage=\beta_0+\delta_0marrmale+\delta_1marrfem+\delta_2singfem+\beta_1educ+\dots+u$ (base group = single men)

2 Interaction involving Dummy Variables

(1) categories. [EG] $Wage=\beta_0+\delta_1female+\delta_2married+\delta_3female\cdot married$

Test $H_0:\delta_3=0$ ⇒ gender differential depends on marital status? vice versa

(2) ordinal information: MBR = municipal binary interest, CR = credit ratings (0-4).

[EG]: $MBR=\beta_0+\beta_1CR+\text{other factors}$ ⇒ constant marginal return to CR

$MBR=\beta_0+\delta_1CR_1+\delta_2CR_2+\delta_3CR_3+\delta_4CR_4+\text{other factors}$

(3) different slopes: $y=\beta_0+\delta_1d_1+\beta_1x+\delta_2d_2x+u$. ⇒ structural change (diff subpopulation)

- Testing for differences Across Group: $y=(\beta_0+\delta_1d_1)+(\beta_1+\delta_2d_2)x+u$

-Method 1: dummy variable approach

unrestricted: $y=\beta_0+\beta_1x_1+\dots\beta_kx_k+\delta_0d+\delta_1d\cdot x_1+\dots\delta_kd\cdot x_k+u$

test the null: $H_0:\delta_0=\delta_1=\dots\delta_k=0$ ($k+1$ restrictions)

restricted: $y=\beta_0+\beta_1x_1+\dots\beta_kx_k+u$

$$F=\frac{(SSR_r-SSR_{ur})/(k+1)}{SSR_{ur}/[n-2(k+1)]}=\frac{(R_{ur}^2-R_r^2)/(k+1)}{(1-R_{ur}^2)/[n-2(k+1)]}\sim F(k+1,n-2(k+1))$$

↙ normality ✓ precise
normality × asymptotic

*HW3: large R^2, large F statistics, low t statistics ⇒ severe multicollinearity

-Method 2: Chew Test / Chow Test.

{ $y=\beta_{0,1}+\beta_{1,1}x_1+\dots\beta_{k,1}x_k+u$ (n_1 observations)
$y=\beta_{0,2}+\beta_{1,2}x_1+\dots\beta_{k,2}x_k+u$ (n_2 observations)

unrestricted model: $n=n_1+n_2$ observations + $2(k+1)$ parameters,
$SSR_{ur}=SSR_1+SSR_2$

test the null: $H_0:\beta_{j,1}=\beta_{j,2}$, for $j=0,1,\dots k$ ($k+1$ restrictions).

restricted: $y=\beta_0+\beta_1x_1+\dots\beta_kx_k+u$ ⇒ obtain SSR_r

$$F=\frac{(SSR_r-(SSR_1+SSR_2))/(k+1)}{(SSR_1+SSR_2)/(n-2(k+1))}\sim F(k+1,n-2(k+1))$$

- can also modify to test only the slope coefficient differences.

$H_0:\delta_1=\dots\delta_k=0$ (k restrictions) — dummy variable approach

$H_0:\beta_{j,1}=\beta_{j,2}$ for $j=1,2,\dots k$ (k restrictions) — Chew approach

restricted: $y=\beta_0+\delta_0d+\beta_1x_1+\dots\beta_kx_k+u$ ⇒ obtain SSR_r.

$$F=\frac{(SSR_r-(SSR_1+SSR_2))/k}{(SSR_1+SSR_2)/(n-2(k+1))}\sim F(k,n-2(k+1))$$

3 Binary Dependent Variables

(1) Linear Probability Model: $y=\beta_0+\beta_1x_1+\beta_2x_2+\dots\beta_kx_k+u$ (LPM)

response probability $p(x)=E(y|\vec x)=P(y=1|\vec x)\cdot1+P(y=0|\vec x)\cdot0=P(y=1|\vec x)\equiv p(\vec x)$.

Remarks: 1. predicted y can be outside [0,1]

2. LPM violates the assumption of homoskedasticity

$Var(y|\vec x)=E(y^2|\vec x)-E^2(y|\vec x)=E(y|\vec x)-[E(y|\vec x)]^2=E(y|\vec x)[1-E(y|\vec x)]$

∴ $Var(y|\vec x)=p(\vec x)[1-p(\vec x)]$

⇒ we need to use heteroskedasticity-robust White standard error

$y_i=G(\beta_0+\beta_1x_{i1}+\dots\beta_kx_{ik})+u_i$
$E(y_i|\vec x)=G(\beta_0+\beta_1x_{i1}+\dots\beta_kx_{ik})=p_i$
∴ $u_i=1-p_i$ (probability $=p_i$) or $-p_i$ $(1-p_i)$

- $P(y=1|x)=G(\beta_0+\beta_1x_1+\dots x_k)=G(\beta_0+\vec x\vec\beta)$

(2) Probit Model (standard normal CDF)

$P(y=1|\vec x)=\Phi(\beta_0+\beta_1x_1+\dots\beta_kx_k)$, where $\Phi(z)=\int_{-\infty}^z\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$

(3) Logit Model (logistic CDF)

$P(y=1|\vec x)=\Lambda(\beta_0+\beta_1x_1+\dots\beta_kx_k)$, where $\Lambda(z)=\frac{e^z}{1+e^z}$

- Derive from Latent Model: $y^*=\beta_0+\beta_1x_1+\dots\beta_kx_k-\varepsilon$. $y=1\{y^*\geq0\}$. where ε has CDF $G(\cdot)$

Assumption: independence between ε and $\vec x=(x_1,x_2,\dots x_k)$

$P(y=1|\vec x)=P(y^*\geq0|\vec x)=P(\varepsilon\leq\beta_0+\beta_1x_1+\dots\beta_kx_k|\vec x)=G(\beta_0+\beta_1x_1+\dots\beta_kx_k)$

…ditional probability distribution for ith observation: $P(y_i|x_i)=p_i^{y_i}(1-p_i)^{1-y_i}$

joint probability distribution of $y_1,\cdots y_n$: $p(y_1,\cdots y_n|\vec{x}_1,\cdots\vec{x}_n)=\prod_{i=1}^n p_i^{y_i}(1-p_i)^{1-y_i}$

← $\mathcal{L}(\beta_0,\beta_1,\cdots\beta_k)=\ln(\prod_{i=1}^n p_i^{y_i}(1-p_i)^{1-y_i})=\sum_{i=1}^n\ln[p_i^{y_i}(1-p_i)^{1-y_i}]=\sum_{i=1}^n\{y_i\ln G(\beta_0+\beta_1x_{i1}+\cdots\beta_kx_{ik})+(1-y_i)\ln G(\beta_0+\beta_1x_{i1}+\cdots\beta_kx_{ik})\}$, where $G(\cdot)=\Phi(\cdot)$ and $\Lambda(\cdot)$ respectively

negative result

Solution: $(\hat\beta_0,\hat\beta_1,\cdots\hat\beta_k)$. (maximize the above log-likelihood function). method: FOCs.

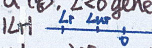
properties: consistent, asymptotically normally distributed, asymptotically efficient

T-test: for single coefficient

$H_0:\beta_{k-q+1}=\cdots\beta_k=0$ (q restrictions) L_{ur}, L_r: log-likelihood value for restricted and unrestricted models

likelihood ration statistic $LR=2(L_{ur}-L_r)$, $LR\xrightarrow{d}\chi^2(q)$, $L<0$ generally

reject the null at α level if $LR>\chi^2_\alpha(q)$, $|L_{ur}|\le|L_r|$



goodness of fit

pseudo $R^2=1-L_{ur}/L_0$ $\in[0,1]$ (L_0: log-likelihood function with only an intercept term)

Interpretation of $\hat\beta$: marginal effect of x_j on the probability of y taking value 1 is given by $g(\beta_0+\beta_1x_1+\cdots\beta_kx_k)\beta_j$. Or: $\Delta\hat{p}(y=1|\vec{x})\approx[g(\hat\beta_0+\hat\beta_1x_1+\cdots\hat\beta_kx_k)\hat\beta_j]\Delta x_j$.

- PEA (partial effect at the average): $g(\hat\beta_0+\hat\beta_1\bar{x}_1+\cdots\hat\beta_k\bar{x}_k)\hat\beta_j$.
- APE (average partial effect): $\frac{1}{n}\sum_{i=1}^n g(\hat\beta_0+\hat\beta_1x_{i1}+\cdots\hat\beta_kx_{ik})\hat\beta_j$.

(*) $P(y=1|\vec{x})=G(\beta_0+\beta_1x_1+\cdots\beta_kx_k)$, $\frac{\partial P(y=1|\vec{x})}{\partial x_j}=\beta_j g(\beta_0+\beta_1x_1+\cdots\beta_kx_k)$ ⇒ marginal effect

⇒ dummy x_k: $\hat{p}(y=1|x_k=1,\overline{other})-\hat{p}(y=1|x_k=0,\overline{other})=G(\hat\beta_0+\hat\beta_1\bar{x}_1+\cdots\hat\beta_k)-G(\hat\beta_0+\hat\beta_1\bar{x}_1+\cdots\hat\beta_{k-1}\bar{x}_{k-1})$

8

■ Multiple Regression Analysis: Heteroskedasticity

(conditionally) heteroskedasticity: $Var(u_i|x_i)=\sigma_i^2=\sigma(x_i)^2$ ⇒
1. OLS: unbiased, consistent
2. $R^2,\bar{R}^2$: consistent of population R^2
3. OLS no longer BLUE / asy. efficient
4. t test, F test, LM test fail

OLS: estimator ✓, reestimate s.e. and adjust test statistics

This method is known as: heteroskedasticity-robust procedure ⇒ asymptotically valid $E(u_i|x_i)=0$

1. Variance (SLR: $y_i=\beta_0+\beta_1x_i+u_i$, $E(u_i|\vec{x})=0$, $Var(u_i|\vec{x})=\sigma_i^2$) $\hat\beta_1=\frac{\sum(x_i-\bar{x})y_i}{\sum(x_i-\bar{x})^2}=\beta_1+\frac{\sum(x_i-\bar{x})u_i}{\sum(x_i-\bar{x})^2}$

$Var(\hat\beta_1|\vec{x})=Var(\frac{\sum(x_i-\bar{x})u_i}{\sum(x_i-\bar{x})^2}|\vec{x})=\frac{1}{SST_x^2}\sum_{i=1}^n(x_i-\bar{x})^2Var(u_i|\vec{x})=\frac{1}{SST_x^2}\sum_{i=1}^n(x_i-\bar{x})^2\sigma_i(x_i)^2$

White's heteroskedasticity-robust s.e. for $\hat\beta_1$:

Or: $\frac{n}{n-k-1}$ →

$se(\hat\beta_1)=\sqrt{\frac{\sum_{i=1}^n(x_i-\bar{x})^2\hat{u}_i^2}{SST_x^2}}$ (SLR)

$se(\hat\beta_j)=\sqrt{\frac{n}{n-k-1}\frac{\sum_{i=1}^n\hat{r}_{ij}^2\hat{u}_i^2}{SSR_j^2}}$ (MLR)

($\hat{r}_{ij}$: reg x_j on others), $SSR_j=\sum_{i=1}^n\hat{r}_{ij}^2$

MLR, $y_i=\beta_0+\beta_1x_{i1}+\cdots\beta_kx_{ik}+u_i$, $E(u_i|\vec{x})=0$, $Var(u_i|\vec{x})=\sigma_i^2$

$\hat\beta_j=\beta_j+\frac{\sum_{i=1}^n\hat{r}_{ij}u_i}{\sum_{i=1}^n\hat{r}_{ij}^2}$, $Var(\hat\beta_j|\vec{x})=Var(\frac{\sum_{i=1}^n\hat{r}_{ij}u_i}{\sum_{i=1}^n\hat{r}_{ij}^2}|\vec{x})=\frac{1}{(\sum_{i=1}^n\hat{r}_{ij}^2)^2}\sum_{i=1}^n\hat{r}_{ij}^2\sigma_i^2$

Thus we have $se(\hat\beta_j)=\sqrt{\sum_{i=1}^n\hat{r}_{ij}^2\hat{u}_i^2/(\sum_{i=1}^n\hat{r}_{ij}^2)^2}=\sqrt{\sum_{i=1}^n\hat{r}_{ij}^2\hat{u}_i^2/SSR_j^2}$

df-correction: $se(\hat\beta_j)=\sqrt{\frac{n}{n-k-1}\sum_{i=1}^n\hat{r}_{ij}^2\hat{u}_i^2/SSR_j^2}$

SLR: $Var(\hat\beta_1|\vec{x})=\frac{\sigma^2}{SST_x}$ (homo), $Var(\hat\beta_1|\vec{x})=\frac{\sum_{i=1}^n(x_i-\bar{x})^2\sigma_i^2}{SST_x^2}$ (hetero)

$se(\hat\beta_1|\vec{x})=\sqrt{\frac{\hat\sigma^2}{SST_x}}$ (homo), $se(\hat\beta_1|\vec{x})=\sqrt{\frac{\sum_{i=1}^n(x_i-\bar{x})^2\hat{u}_i^2}{SST_x^2}}$ (hetero).

large sample size (n): ☆
- heteroskedasticity-robust s.e.
- robust t statistics,
- robust F statistics = Wald statistics.

MLR: $Var(\hat\beta_j|\vec{x})=\frac{\sigma^2}{SST_j(1-R_j^2)}=\frac{\sigma^2}{SSR_j}=\frac{\sigma^2}{\sum_{i=1}^n\hat{r}_{ij}^2}$ (homo), $Var(\hat\beta_j|\vec{x})=\frac{\sum_{i=1}^n\hat{r}_{ij}^2\sigma_i^2}{SSR_j^2}$ (hetero) $SSR_j=\sum_{i=1}^n\hat{r}_{ij}^2$

$se(\hat\beta_j|\vec{x})=\sqrt{\frac{\hat\sigma^2}{SSR_j}}=\sqrt{\frac{\hat\sigma^2}{\sum_{i=1}^n\hat{r}_{ij}^2}}$ (homo), $se(\hat\beta_j|\vec{x})=\sqrt{\frac{\sum_{i=1}^n\hat{r}_{ij}^2\hat{u}_i^2}{SSR_j^2}}$ (hetero) robust

(*) Robust LM Statistic: 1. reg on restricted (q) model → $\check{u}$ 2. q regressions on included variable → $\check{r}_1,\cdots\check{r}_q$ 3. reg 1 on $\check{r}_1\check{u},\check{r}_2\check{u},\cdots\check{r}_q\check{u}$ without intercept 4. $LM=n-SSR_1\sim\chi^2(q)$

2 Test for heteroskedasticity $H_0: Var(u|x_1,\cdots x_k)=\sigma^2 \Leftrightarrow H_0: E(u^2|x_1,\cdots x_k)=E(u^2)=\sigma^2$ (MLR1-4)

- **Breusch-Pagan Test.** $u^2=\delta_0+\delta_1x_1+\cdots\delta_kx_k+v$, testing $H_0:\delta_0=\delta_1=\cdots\delta_k=0$
  1. reg y_i on $(1,x_{i1},\cdots x_{ik})$ ⇒ residual $\hat{u}_i$
  2. reg $\hat{u}_i^2$ on $(1,x_{i1},\cdots x_{ik})$ ⇒ the $R^2_{\hat{u}_i^2}$
  3. $F=\frac{R^2_{\hat{u}^2}/k}{(1-R^2_{\hat{u}^2})/(n-k-1)}\sim F(k,n-k-1)$. or $LM=nR^2_{\hat{u}^2}\xrightarrow{d}\chi^2(k)$ (asymptotically).

  (*) BP test only detects any linear forms of hetereoskedastiaty
- **White Test** can detect nonlinear forms of heteroskedasticity
  1. reg y_i on $(1,x_{i1},\cdots x_{ik})$ ⇒ residual $\hat{u}_i$
  2. reg $\hat{u}_i^2$ on $(1,x_{i1},\cdots x_{ik},x_{i1}^2,\cdots x_{ik}^2,x_{i1}\cdot x_{ih})$ ⇒ $R^2_{\hat{u}_i^2}$
  3. $LM=nR^2_{\hat{u}_i^2}\xrightarrow{d}\chi^2(p)$, p = N(regressors in "2") (×1)

  Remarks:
  1. p = non-constant, distinct regressors
  2. Special form of White Test: reg $\hat{u}_i^2$ on $(1,\hat{y}_i,\hat{y}_i^2)$ ⇒ $R^2_{\hat{u}_i^2}$ (asy.) $LM=nR^2\sim\chi^2(2)$, reject when $LM>\chi^2_\alpha(2)$

3. Weighted Least Squares (WLS) and Generalized Least Squares (GLS)

GLS: transform to models with homoskedastic errors ⇒ new t and F asymptotic distribution

Suppose: $Var(u|\vec{x})=\sigma^2h(\vec{x})$, $h_i=h(x_{i1},\cdots x_{ik})>0$.

$y_i=\beta_0+\beta_1x_{i1}+\cdots\beta_kx_{ik}+u_i$ ⇒ $\frac{y_i}{\sqrt{h_i}}=\beta_0\frac{1}{\sqrt{h_i}}+\beta_1\frac{x_{i1}}{\sqrt{h_i}}+\cdots\beta_k\frac{x_{ik}}{\sqrt{h_i}}+\frac{u_i}{\sqrt{h_i}}$

define $y_i^*=\frac{y_i}{\sqrt{h_i}}$, $x_{ij}^*=\frac{x_{ij}}{\sqrt{h_i}}$ ⇒ $y_i^*=\beta_0x_{i0}^*+\beta_1x_{i1}^*+\cdots\beta_kx_{ik}^*+u_i^*$, $(u_i^*=\frac{u_i}{\sqrt{h_i}})$.

Then we have $Var(u_i^*|\vec{x}_i)=\frac{Var(u_i|\vec{x}_i)}{h_i}=\sigma^2$ ⇒ homoskedastic

Theorem: GLS is BLUE; t and F statistics are valid; R^2 not informative (has no meaning in explaining sample variation of y.)

GLS is a special case of WLS: obtain GLS by choosing $\vec\beta_j$ to minimize the WLS objective function:

$w_i=\frac{1}{h_i}$ $\sum_{i=1}^n(y_i-\beta_0-\beta_1x_{i1}-\cdots\beta_kx_{ik})^2/h_i$ ($\frac{1}{h_i}$ is the weight) ⇔ $\sum_{i=1}^n(y_i^*-\beta_0x_{i0}^*-\beta_1x_{i1}^*-\cdots\beta_kx_{ik}^*)^2$ $x_{ij}^*=\frac{x_{ij}}{\sqrt{h_i}}$

A WLS estimator can be defined for any positive weights $w_i>0$: $\sum_{i=1}^n(y_i-\beta_0-\cdots\beta_kx_{ik})^2w_i$ $(x_{i0}^*=1)$

4. Feasible GLS (unknown conditional variances) — FGLS

Assumption: $Var(u|\vec{x})=\sigma^2\exp(\delta_0+\delta_1x_1+\cdots\delta_kx_k)$ ⇒ $u^2=\sigma^2\exp(\delta_0+\delta_1x_1+\cdots\delta_kx_k)v$ $(E(v|\vec{x})=1)$

⇒ $\log(u^2)=\alpha_0+\delta_1x_1+\delta_2x_2+\cdots\delta_kx_k+e$, where $\alpha_0=\log(\sigma^2)+\delta_0$, $e=\log v$

run $\log(\hat{u}^2)$ on $(1,x_{i1},\cdots x_{ik})$ ⇒ OLS estimators $\hat\alpha_0,\hat\delta_1,\hat\delta_2,\cdots\hat\delta_k$, fitted value $\hat{g}_i$

Theorem: we can estimate $h_i=h_i(\vec{x})$ by $\hat{h}_i=\exp(\hat{g}_i)$. and use WLS weight $w_i=1/\hat{h}_i$

Remarks: not unbiased in finit sample, consistent, asymptotically more efficient than OLS. (not BLUE)

missspecified $h_i(x)$ ⇒ If strong heteroskedasticity ⇒ WLS (generally consistent, but WLS s.e. and test statistics may wrong) make valid inference ⇒ White's heteroskedasticity-robust s.e. with OLS unless specific h_i ⇒ FGLS

5. LPM: OLS with White's s.e. (but not that efficient): WLS method. problem: when $\hat{y}_i<0$ or $\hat{y}_i>1$ ⇒ $\hat{h}_i<0$ need adjustment

estimate $h_i=p(\vec{x}_i)[1-p(\vec{x}_i)]$ by $\hat{h}_i=\hat{y}_i(1-\hat{y}_i)$, with $\hat{y}_i=\hat\beta_0+\hat\beta_1x_{i1}+\cdots\hat\beta_kx_k$ (OLS estimators)

use $\hat{h}_i$ to obtain FGLS estimators, or use $1/\hat{h}_i$ as weights to obtain WLS estimator.

9

■ Multiple Regression Analysis: More on Specification and Data Issues

1. Test of Functional form (the misspecification problem)

(1) Test for omitted nonlinearity. **F-test / Wald test, LM test**

Often chose: $z_{1i}=x_i^2$, $z_{2i}=x_i^3$, m = 1, 2 or 3

SLR: $y_i=\beta_0+\beta_1x_{i1}+u_i$. let $z_{ji}=h_j(x_i)$ denotes nonlinear function of x_i for j = 1, 2, ⋯ m.

reg: $y_i=\alpha_0+\alpha_1x_i+\gamma_1z_{1i}+\cdots\gamma_mz_{mi}+v_i$ (OLS approach)

$H_0:\gamma_1=\gamma_2=\cdots\gamma_m=0$. using F test or LM test

(2) **Ramsey's RESET** (REgression Specification Error Test)
1. reg. $y_i=\beta_0+\beta_1x_{i1}+\cdots\beta_kx_{ik}+u_i$ ⇒ fitted value $\hat{y}_i$
2. reg. auxillary regression: $y_i=\beta_0+\beta_1x_{i1}+\cdots\beta_kx_{ik}+\gamma_1\hat{y}_i^2+\gamma_2\hat{y}_i^3+\cdots\gamma_{m-1}\hat{y}_i^m+\varepsilon_i$
3. Form F or LM statistic for testing $H_0:\gamma_1=\gamma_2=\cdots\gamma_{m-1}=0$ $F\sim F(m-1,n-k-m)$, $LM\sim\chi^2(m-1)$. (asymptotically)
4. reject: $F>F_\alpha(m-1,n-k-m)$ or $LM>\chi^2_\alpha(m-1)$. m often chooses 2, 3, 4.

2. Proxy for Unobserved Explanatory Variables

- True model: $y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3^*+u$. (e.g. x_3^*: ability) — latent / unobserved

  proxy variable: $x_3^*=\delta_0+\delta_3x_3+v_3$. $E(v_3|x_3)=0$

  Regression: $y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3(\delta_0+\delta_3x_3+v_3)+u=(\beta_0+\beta_3\delta_0)+\beta_1x_1+\beta_2x_2+\beta_3\delta_3x_3+(u+\beta_3v_3)$.

2. If $E(x_3^*|x_1,x_2,x_3)=\delta_0+\delta_1x_1+\delta_2x_2+\delta_3x_3$ ⇒ regression: $y=(\beta_0+\delta_0)+(\beta_1+\beta_3\delta_1)x_1+(\beta_2+\beta_3\delta_2)x_2+\beta_3\delta_3x_3+(u+\beta_3v_3)$.

↑ unless $\delta_1=\delta_2=0$ β_1,β_2 will be biased

Classical Conditions: 1. u is uncorrelated with x_1,x_2,x_3^*,x_3.

2. v_3 is uncorrelated with x_1,x_2,x_3. → (stronger): $E(x_3^*|x_1,x_2,x_3)=E(x_3^*|x_3)$ $=\delta_0+\delta_3x_3$

Consequence: 1. problem of collinearity (for unbiasedness we must ignore) ⇒ $Var(\hat\beta)$ not necessarily boosted.

- Lagged Dependent Variable: [e.g.] $crime=\beta_0+\beta_1unem+\beta_2expend+\beta_3crime_{-1}+u$

3. Measurement Error

(1) Error in dependent variable

True model: $y^*=\beta_0+\beta_1x_1+\cdots\beta_kx_k+u$

observe $y=y^*+e_0$ (measurement error: e_0)

1. $E(e_0)=0$. 2. e_0 is uncorrelated with $y^*,x_1,x_2,\cdots x_k$. $Var(u+e_0)=Var(u)+Var(e_0)>Var(u)$ in general

Regression: $y=\beta_0+\beta_1x_1+\cdots\beta_kx_k+u+e_0$. ⇒ unbiased, consistent, larger variance

↑ notice: $Cov(y^*,e_0)=0$ is not necessary for consistency

(2) Error in Independent Variables

True model: $y_i=\beta_0+\beta_1x_{i1}^*+u_i$. But x_{i1}^* is measured with error: $x_{i1}=x_{i1}^*+e_{i1}$

Assumption: $E(e_{i1})=0$. $E(y_i|x_{i1}^*,x_{i1})=E(y_i|x_{i1}^*)=\beta_0+\beta_1x_{i1}^*$ — But $Cov(e_{i1},x_{i1})$? $Cov(e_{i1},x_{i1}^*)$?

Regression: $y_i=\beta_0+\beta_1x_{i1}+(u-\beta_1e_{i1})$ — $(u-\beta_1e_{i1})$ is correlated with x_{i1}

- $Cov(e_{i1},x_{i1})=0$. (unlikely) —— consistent OLS estimator but with a larger variance

  ⇒ $E(u_i-\beta_1e_{i1})=0$, $Var(u_i-\beta_1e_{i1})=Var(u_i)+\beta_1^2Var(e_{i1})\equiv\sigma_u^2+\beta_1^2\sigma_{e_1}^2$ analogue to proxy conditions
- $Cov(e_{i1},x_{i1}^*)=0$ (most likely)

  CEV: 1. $E(e_{i1})=Cov(e_{i1},x_{i1}^*)=0$. 2. u_i is uncorrelated with (e_{i1},x_{i1}^*): $E(e_{i1}u_i)=E(x_{i1}^*u_i)=0$ ⇒ $E(u_i|x_{i1})=0$

  ⇒ $Cov(e_{i1},x_{i1})=Cov(e_{i1},x_{i1}^*+e_{i1})=\sigma_{e_1}^2$ denotes $\sigma_{x_1^*}^2=Var(x_{i1}^*)$

  $Cov(u-\beta_1e_{i1},x_{i1})=-\beta_1Cov(e_{i1},x_{i1})=-\beta_1\sigma_{e_1}^2$

  ⇒ $\hat\beta_1=\beta_1+\frac{\sum_{i=1}^n(x_{i1}-\bar{x}_1)(u_i-\beta_1e_{i1})}{\sum_{i=1}^n(x_{i1}-\bar{x})^2}\xrightarrow{P}\beta_1+\frac{Cov(x_{i1},u_i-\beta_1e_{i1})}{Var(x_{i1})}=\beta_1-\beta_1\frac{\sigma_{e_1}^2}{\sigma_{x_1}^2}=\beta_1\frac{\sigma_{x_1^*}^2}{\sigma_{x_1^*}^2+\sigma_{e_1}^2}$

  attenuation bias ($|\hat\beta_1|<|\beta_1|$).

  MLR: $y_i=\beta_0+\beta_1x_{i1}^*+\beta_2x_{i2}+\cdots\beta_kx_{ik}+u_i$, $x_{i1}=x_{i1}^*+e_{i1}$

  Assumptions: $E(e_{i1})=0$, u_i is uncorrelated with $x_{i1}^*,\cdots x_{ik}$

  Regression: $y_i=\beta_0+\beta_1x_{i1}+\cdots\beta_kx_{ik}+(u_i-\beta_1e_{i1})$

  ⇒ $\hat\beta_1=\beta_1+\frac{\sum_{i=1}^n\hat{r}_{i1}(u_i-\beta_1e_{i1})}{\sum_{i=1}^n\hat{r}_{i1}^2}$, $plim(\hat\beta_1)=\beta_1\frac{\sigma_{r_1}^2}{\sigma_{r_1}^2+\sigma_{e_1}^2}$ ⇒ attenuation bias

  where $\sigma_{r_1}^2$ is the variance of the error term r_{i1} in reg x_{i1} on $(1,x_{i2},\cdots x_{ik})$.

Measurement error in single variable causes inconsistency in other $\hat\beta_j$ too, unless $Cov(x_j,x_{i1}^*)=0$

4. Missing Data, Nonrandom Samples and Outlying Observations

MCAR (missing completely at random): no affect (sample restricted to observation).

— missing is independent of both the observed and unobserved factor (x and u)

Missing systematically: violate MLR. 2

(1) exogenous sample selection (chosed based on x) (e.g.) $saving=\beta_0+\beta_1income+\beta_2age+u$.

— unbiased, consistent estimators (provided enough variation) → inefficient

(2) endogenous sample selection (chosed based on y). (e.g.) $wealth=\beta_0+\beta_1educ+\beta_2exper+u$.
- truncated regression model: $E(y|y<a,x)\ne E(y|x)$.
- sample selection model: $E(y|y\text{ is observed},x)\ne E(y|x)$

Outliers: large effect & different population indication, may provide important info. in Var.

— Least Absolute Deviations (LAD): a method robust to outliers

$\min_{(\beta_0,\beta_1,\cdots\beta_k)}\sum_{i=1}^n|y_i-\beta_0-\beta_1x_{i1}-\cdots-\beta_kx_{ik}|$ ⇒ $median(y|\vec{x})=\beta_0+\beta_1x_1+\cdots\beta_kx_k$ ($median(u|\vec{x})=0$). quantile regression

⇒ estimate the effect of the x_j on different parts of the distribution y.

■ Chapter 15 Instrumental Variables and 2SLS

$$\begin{tikzcd} z \arrow[r] & x \arrow[r] & y \\ & u \arrow[u] \arrow[ur] & \end{tikzcd}$$

Endogenity: the explanatory variable is correlated with the error term in structural eq.

1. OVB: $y_i=\beta_0+\beta_1x_{i1}+v_i$, $v_i=\beta_2x_{i2}+u_i$ ⇒ $Cov(x_{i1},v_i)=Cov(x_{i1},\beta_2x_{i2}+u_i)=\beta_2Cov(x_{i1},x_{i2})$ ⇒ $\hat\beta_1=\beta_1+\beta_2\frac{Cov(x_{i1},x_{i2})}{Var(x_{i1})}$ (asymptotic)
2. Measurement bias: $y_i=\beta_0+\beta_1x_i+v_i$, $v_i=u_i-\beta_1e_i$ $(x_i=x_i^*+e_i)$ ⇒ $Cov(x_i,v_i)=-\beta_1\sigma_{e_i}^2$ ⇒ $\hat\beta_1=\beta_1-\beta_1\frac{\sigma_{e_i}^2}{\sigma_x^2}=\beta_1\frac{\sigma_{x^*}^2}{\sigma_{x^*}^2+\sigma_{e_i}^2}$ (asymptotic)
3. Simultaneous Equation Bias ($E(e_{i1})=E(e_{i2})=0$, $Var(e_{i1})=\sigma_1^2$; $Var(e_{i2})=\sigma_2^2$; $E(e_{i1}e_{i2})=0$)

   $\begin{cases}\text{Demand}: q_i=\beta_0+\beta_1p_i+e_{1i}\\ \text{Supply}: q_i=\gamma_0+\gamma_1p_i+e_{2i}\end{cases}$ ⇒ $\begin{cases}p_i=\frac{1}{\beta_1-\gamma_1}(\gamma_0-\beta_0+e_{2i}-e_{1i})\\ q_i=\frac{1}{\beta_1-\gamma_1}(\beta_1\gamma_0-\beta_0\gamma_1+\beta_1e_{2i}-\gamma_1e_{1i})\end{cases}$ (*) $\beta_1<0$, $\gamma_1>0$

   $Cov(e_{1i},p_i)=\frac{-1}{\beta_1-\gamma_1}\sigma_1^2>0$ ⇒ $\hat\beta_1=\beta_1+\frac{Cov(p_i,e_{1i})}{Var(p_i)}=\beta_1-\frac{\sigma_1^2}{\beta_1-\gamma_1}\ne\beta_1$ (asymptotic)
4. Sample Selection Bias. use inverse Mill ratio from the truncated distribution

1. Instrumental Variables (IV) in SLR

FOCs:
$\begin{cases}\frac{1}{n}\sum_{i=1}^n(y_i-\hat\beta_0-\hat\beta_1x_i)=0\\ \frac{1}{n}\sum_{i=1}^n z_i(y_i-\hat\beta_0-\hat\beta_1x_i)=0\end{cases}$ ↓ $y_i=\beta_0+\beta_1x_i+u_i$ $E(u_i)=E(z_iu_i)=0$

Conditions: 1. Instrument exgeneity: $Cov(z,u)=0$. ← exclusion condition 2. Instrument Relevance: $Cov(z,x)\ne0$. $Cov(x,u)\ne0$

⇒ $\hat\beta_1=\frac{\sum_{i=1}^n(z_i-\bar{z})y_i}{\sum_{i=1}^n(z_i-\bar{z})x_i}\xrightarrow{P}\beta_1=\frac{Cov(z_i,y_i)}{Cov(z_i,x_i)}$ $(\hat\beta_1\xrightarrow{P}\frac{\frac{1}{n}\sum_{i=1}^n(z_i-\bar{z})u_i}{\frac{1}{n}\sum_{i=1}^n(z_i-\bar{z})x_i}+\beta_1=\beta_1+\frac{Cov(z,u)}{Cov(z,x)}=\beta_1)$

$\frac{\frac{1}{n}\sum_{i=1}^n(z_i-\bar{z})y_i}{\frac{1}{n}\sum_{i=1}^n(z_i-\bar{z})x_i}$

Assume homoskedasticity $(E(u^2|z)=\sigma^2)$: $\sqrt{n}(\hat\beta_{1,IV}-\beta_1)\xrightarrow{d}N(0,\frac{\sigma_z^2\sigma^2}{Cov(x,z)^2})=N(0,\frac{\sigma^2}{\sigma_x^2\rho_{xz}^2})$

comparison: $\sqrt{n}(\hat\beta_{1,OLS}-\beta_1)\xrightarrow{d}N(0,\frac{\sigma^2}{\sigma_x^2})$.

IV estimator is less efficient than OLS.

Note that $\sqrt{n}(\hat\beta_1-\beta_1)=\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n(z_i-\bar{z})u_i}{\frac{1}{n}\sum_{i=1}^n(z_i-\bar{z})x_i}$, denominator $\xrightarrow{P}Cov(x,z)$ (LLN).

numerator $=\frac{1}{\sqrt{n}}\sum_{i=1}^n(z_i-\bar{z})u_i=\frac{1}{\sqrt{n}}\sum_{i=1}^n[z_i-E(z_i)]u_i+[E(z_i)-\bar{z}]\frac{1}{\sqrt{n}}\sum_{i=1}^nu_i$ (converge to 0 by probability)

where $N(0,Var([z-E(z)]u))=N(0,E[(z-E(z))^2u^2])=N(0,E([z-E(z)]^2E(u^2|z))=N(0,\sigma_z^2\sigma^2)$

Thus $\sqrt{n}(\hat\beta_1-\beta_1)\xrightarrow{d}\frac{N(0,\sigma_z^2\sigma^2)}{Cov(x,z)}=N(0,\frac{\sigma^2}{\sigma_x^2\rho_{xz}^2})$. where $\rho_{xz}=\frac{Cov(z,x)}{\sigma_x\cdot\sigma_z}$

Recall: $\hat\sigma^2=\frac{\sum_{i=1}^n\hat{u}_i^2}{n-k-1}$ $se(\hat\beta_1)=\frac{\hat\sigma}{\sqrt{SST_x}}$

IV residual $\hat{u}_i=y_i-\hat\beta_0-\hat\beta_1x_i$, notes that $Var(\hat\beta_1)=\frac{\sigma^2}{n\sigma_x^2\rho_{xz}^2}$ (asy.)

⇒ estimate Asy-sd($\hat\beta_1$) by: $se(\hat\beta_1)=\sqrt{\frac{\hat\sigma^2}{SST_xR_{xz}^2}}$, where $\hat\sigma^2=\frac{1}{n-2}\sum_{i=1}^n\hat{u}_i^2$, $SST_x=\sum_{i=1}^n(x_i-\bar{x})^2$ R_{xz}: sample correlation (x, z)

asymptotically unbiased and consistent

(*) 1. IV estimator boast $se(\hat\beta_1)$, often making asy. s.e much larger ↘ $Var(\hat\beta_1)=\frac{\sigma^2}{n\sigma_x^2\rho_{xz}^2}$

2. Instrumental Variables (IV) in MLR (interested in casual relationship) ↓ reduced form equation

structural Equation: $y_1=\beta_0+\beta_1y_2+\beta_2z_1+u_1$. ($z_2$ is the excluded instrument for y_2: $Cov(z_2,u_1)=0$.

RF → $y_2=\pi_0+\pi_1z_1+\pi_2z_2+v_2$, $\pi_2\ne0$ ⇔ $Cov(z_2,y_2)\ne0$).

Moment Conditions:

$\begin{cases}E(u_1)=0\\ Cov(z_1,u_1)=0 \Rightarrow E(z_1u_1)=0\\ Cov(z_2,u_1)=0 \Rightarrow E(z_2u_1)=0\end{cases}$ ⇒ $\begin{cases}\sum_{i=1}^n(y_{i1}-\hat\beta_0-\hat\beta_1y_{i2}-\hat\beta_2z_{i1})=0\\ \sum_{i=1}^n z_{i1}(y_{i1}-\hat\beta_0-\hat\beta_1y_{i2}-\hat\beta_2z_{i1})=0\\ \sum_{i=1}^n z_{i2}(y_{i1}-\hat\beta_0-\hat\beta_1y_{i2}-\hat\beta_2z_{i1})=0\end{cases}$ ⇒ asymptotically unbiased and normally distributed (differ from OLS)

3. Two Stage Least Squares (2SLS) Estimation consistent, asy-unbiased, normally distributed

structural equation: $y_1=\beta_0+\beta_1y_2+\beta_2z_1+u_1$. excluded IVs: z_2,z_3

reduced form (RF) equation: $y_2=\pi_0+\pi_1z_1+\pi_2z_2+\pi_3z_3+v_2=y_2^{clean}+y_2^{dirty}$, $Cov(v_2,u)\ne0$.

⇒ the best IV for y_2: $y_2^*=y_2^{clean}=\pi_0+\pi_1z_1+\pi_2z_2+\pi_3z_3$ ($\pi_2\ne0$ or $\pi_3\ne0$).

1. run reg y_{i2} on $(1,z_{i1},z_{i2},z_{i3})$: $\hat{y}_{i2}=\hat\pi_0+\hat\pi_1z_{i1}+\hat\pi_2z_{i2}+\hat\pi_3z_{i3}$
2. run reg y_{i1} on $(1,\hat{y}_{i2},z_{i1})$ to obtain estimates $\hat\beta_j$'s. ⇒ $\begin{cases}\sum_{i=1}^n(y_{i1}-\hat\beta_0-\hat\beta_1\hat{y}_{i2}-\hat\beta_2z_{i1})=0\\ \sum_{i=1}^n z_{i1}(y_{i1}-\hat\beta_0-\hat\beta_1\hat{y}_{i2}-\hat\beta_2z_{i1})=0\\ \sum_{i=1}^n\hat{y}_{i2}(y_{i1}-\hat\beta_0-\hat\beta_1\hat{y}_{i2}-\hat\beta_2z_{i1})=0\end{cases}$

Multiple Endogenous Explanatory Variables over-identifying restrictions: m−k

m<k: underidentified m=k: exactly ~

structural model: $y_1=\beta_0+\beta_1y_2+\beta_2y_3+\cdots\beta_ky_{k+1}+\beta_{k+1}z_1+\cdots\beta_{k+r}z_r+u_1$ k endogenous y, r exogenous z, m IVs z

excluded IVs: $z_{r+1},\cdots z_{r+m}$ (m≥k) exactly / overidentified

⇒ 1. j = 2, 3, ⋯ k+1, reg y_{ij} on $(1,z_{i1},\cdots z_{i,r+m})$: $\hat{y}_{ij}=\hat\pi_{j0}+\hat\pi_{j1}z_{i1}+\cdots\hat\pi_{j,m+r}z_{i,r+m}$

2. run reg y_{i1} on $(1,\hat{y}_{i2},\cdots\hat{y}_{i,k+1},z_{i1},\cdots z_{ir})$ to obtain estimates $\hat\beta_j$'s.

(*) bigger k: reduce asy. variance but tends to increase finit sample bias

4. Instrument Relevance (eg. single endogenous 2SLS). $y_1=\beta_0+\beta_1y_2+\beta_2z_1+\cdots\beta_{r+1}z_r+u_1$. IVs $(z_{r+1},\cdots z_m)$
- weak IVs: (finit sample) coefficient bias, asymptotic distribution biased. (✗ asy. normal).

  (SLR): $\hat\beta_1=\frac{S_{zy}}{S_{zx}}$ ⇒ small S_{zx} leads to bad $\hat\beta_{1,IV}$ distribution ($\pi_{r+1},\cdots\pi_{r+m}$ are nearly 0).
- $Cov(z,u)\ne0$. $\begin{cases}IV: plim\hat\beta_{IV}-\beta_1=\frac{Corr(z,u)}{Corr(z,x)}\cdot\frac{\sigma_u}{\sigma_x}\\ OLS: plim\hat\beta_{OLS}-\beta_1=Corr(x,u)\cdot\frac{\sigma_u}{\sigma_x}\end{cases}$ prefer IV to OLS: $\frac{Corr(z,u)}{Corr(z,x)}<Corr(x,u)$. ← $E(u^2|z)=\sigma_u^2$ $E(u^2|x)=\sigma_u^2$

one endogenous regressor ↳ **Test for Instrument Relevance**:

rule of thumb: F>10 (larger F, better IVs).

$H_0:\pi_{r+1}=\pi_{r+2}=\cdots\pi_{r+m}=0$ or nearly 0

null hypothesis:
1. bias of 2SLS less than given
2. actual size of nomial 5% 2SLS t-test ≤ 15% different from H_0^*: $\pi_{r+1}=\cdots\pi_{r+m}=0$

5. **Testing for Over-identifying Restrictions** (no valid test when m≤k). $\hat\beta_j$ by 2SLS ↓

When m=k, $\frac{1}{n}\sum_{i=1}^n\hat{u}_iz_{ij}=0$ for j = 1, 2, ⋯ r+m ⇒ no way to check if $E(u_iz_{ij})=0$

Step 1. run the structural equation by 2SLS ⇒ obtain residual $\hat{u}_1=y_1-\hat\beta_0-\hat\beta_1y_2-\cdots\hat\beta_ky_{k+1}-\hat\beta_{k+r}z_r$

step 2. reg. $\hat{u}_1=\delta_0+\delta_1z_1+\delta_2z_2+\cdots\delta_rz_r+\delta_{r+1}z_{r+1}+\cdots\delta_{r+m}z_{r+m}+e_1$ obtain R_1^2

$\delta_{r+1}=0$ $\delta_{r+2}=0$ $\cdots\delta_{r+m}=0$ ↙
- H_0: all instruments exogenous. J statistic: $J=m\cdot F\overset{asy}{\sim}\chi^2(m-k)$. reject: $J>\chi^2_\alpha(m-k)$ Then At least some instruments are endogenous.
- if homoskedasticity: $nR_1^2\xrightarrow{d}\chi^2(m-k)$. reject: $nR_1^2>\chi^2_\alpha(m-k)$

6. **Testing for Endogenerity** (structural equation: $y_1=\beta_0+\beta_1y_2+\beta_2z_1+\beta_3z_2+u_1$, IVs: z_3,z_4).

RF: $y_2=\pi_0+\pi_1z_1+\pi_2z_2+\pi_3z_3+\pi_4z_4+v_2$. (If y_2 is endogenous: $Cov(v_2,u_1)\ne0$).

⇒ $u_1=\delta_1v_2+e_1$. $Cov(e_1,v_2)=0$, $E(e_1)=0$ ⇒ $H_0:\delta_1=0$ ($\delta_1=0$ ⇒ $Cov(v_2,u_1=0)$).

- If v_2 is observed. $y_1=\beta_0+\beta_1y_2+\beta_2z_1+\beta_3z_2+\delta_1v_2+e_1$ and test $H_0:\delta_1=0$
- Else: Hausman Test.

  single

  step 1. run RF ⇒ obtain $\hat{v}_2$

  step 2. run $y_1=\beta_0+\beta_1y_2+\beta_2z_1+\beta_3z_2+\delta_1\hat{v}_2+e_1^*$

  step 3. Test $H_0:\delta_1=0$.

  1. RF: $y_2=\pi_0+\pi_1z_1+\cdots\pi_4z_4+v_2$ ⇒ obtain $\hat{v}_2$.

  run: $y_1=\beta_0+\beta_1y_2+\beta_2z_1+\beta_3z_2+\delta\hat{v}_2+e_1^*$

  Test $H_0:\delta=0$

(*) Multiple Explanatory Variables:

Test $H_0:\delta_1=\delta_2=\cdots\delta_k=0$

$y_1=\beta_0+\beta_1y_2+\cdots\beta_ky_{k+1}+\beta_{k+1}z_1+\cdots\beta_{k+r}z_r+u_1$

Step 1. RF ⇒ $\hat{v}_j$

Step 2. $y_1=\beta_0+\cdots+\beta_{k+r}z_r+\delta_1\hat{v}_2+\cdots\delta_k\hat{v}_{k+1}+e_1^*$

Step 3. test $H_0:\delta_1=\cdots\delta_k=0$