Econometric Analysis of Cross Section and Panel Data

Lecture 3: Finite-Sample Properties of the OLS Estimator

Zhian Hu

Central University of Finance and Economics

Fall 2024

This Lecture

- Hansen (2022): Chapter 4
- We investigate some finite-sample properties of the least squares estimator in the linear regression model.
- In particular we calculate its finite-sample expectation and covariance matrix and propose standard errors for the coefficient estimators.

Sample Mean in an Intercept-Only Model

有限样本性质　OLS所得为\frac{\sum Y_i}{n}

- The intercept-only model: $Y = \mu + e$, assuming that $\mathbb{E}[e] = 0$ and $\mathbb{E}[e^2] = \sigma^2$.
- In this model $\mu = \mathbb{E}[Y]$ is the expectation of Y. Given a **random sample**, the least squares estimator $\widehat{\mu} = \bar{Y}$ equals the sample mean.
- We now calculate the expectation and variance of the estimator $\bar{Y}$.

抽球实验Y作为抽球颜色

$$\mathbb{E}[\bar{Y}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} Y_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[Y_i\right] = \mu$$

- An estimator with the property that its expectation equals the parameter it is estimating is called unbiased.
- **An estimator $\widehat{\theta}$ for θ is unbiased if $\mathbb{E}[\widehat{\theta}] = \theta$**

无偏性

OLS不能用E[],因为已知Y。若以\beta形式代入则视n^{-1}\sum Y_i

Sample Mean in an Intercept-Only Model

- We next calculate the variance of the estimator.
- Making the substitution $Y_i = \mu + e_i$ we find $\bar{Y} - \mu = \frac{1}{n}\sum_{i=1}^{n} e_i$.

$$\operatorname{var}[\bar{Y}] = \mathbb{E}\left[(\bar{Y}-\mu)^2\right] = \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} e_i\right)\left(\frac{1}{n}\sum_{j=1}^{n} e_j\right)\right]$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\left[e_i e_j\right] = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2$$

存疑，应为i，i否

$$= \frac{1}{n}\sigma^2.$$

E[e_i]=0,E[e_j]=0

- The second-to-last inequality is because $\mathbb{E}\left[e_i e_j\right] = \sigma^2$ for $i = j$ but $\mathbb{E}\left[e_i e_j\right] = 0$ for $i \neq j$ due to independence.

此处写错，应为E[e_ie_i]=0

Linear Regression Model

- We now consider the **linear regression model**.
- The variables (Y, X) satisfy the linear regression equation

$$
\begin{aligned}
Y &= X'\beta + e \qquad \text{线性} \\
\mathbb{E}[e \mid X] &= 0 \qquad \text{零条件期望}
\end{aligned}
$$

  The variables have finite second moments

$$
\begin{aligned}
\mathbb{E}\left[Y^2\right] &< \infty \\
\mathbb{E}\|X\|^2 &< \infty
\end{aligned}
$$

  and an invertible design matrix $\boldsymbol{Q}_{XX} = \mathbb{E}\left[XX'\right] > 0$ 不完全共线，满秩
- Homoskedastic Linear Regression Model: $\mathbb{E}\left[e^2 \mid X\right] = \sigma^2(X) = \sigma^2$ is independent of X.

Expectation of Least Squares Estimator

- **The OLS estimator is unbiased in the linear regression model.**
- In summation notation:

$$
\begin{aligned}
\mathbb{E}\left[\widehat{\beta} \mid X_1, \ldots, X_n\right] &= \mathbb{E}\left[\left(\sum_{i=1}^{n} X_i X_i'\right)^{-1}\left(\sum_{i=1}^{n} X_i Y_i\right) \mid X_1, \ldots, X_n\right] \\
&= \left(\sum_{i=1}^{n} X_i X_i'\right)^{-1} \mathbb{E}\left[\left(\sum_{i=1}^{n} X_i Y_i\right) \mid X_1, \ldots, X_n\right] \\
&= \left(\sum_{i=1}^{n} X_i X_i'\right)^{-1} \sum_{i=1}^{n} \mathbb{E}\left[X_i Y_i \mid X_1, \ldots, X_n\right] \\
&= \left(\sum_{i=1}^{n} X_i X_i'\right)^{-1} \sum_{i=1}^{n} X_i \mathbb{E}\left[Y_i \mid X_i\right] \\
&= \left(\sum_{i=1}^{n} X_i X_i'\right)^{-1} \sum_{i=1}^{n} X_i X_i' \beta = \beta.
\end{aligned}
$$

X皆已知，故可外乘

X_i和X_j独立

使用到零条件期望E[Y_i|X_i]=X_i\beta

如果存在内生性，就不具有无偏性

Expectation of Least Squares Estimator

- In matrix notation:

$$\begin{aligned}
\mathbb{E}[\widehat{\beta} \mid \boldsymbol{X}] &= \mathbb{E}\left[\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Y} \mid \boldsymbol{X}\right] \\
&= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\mathbb{E}[\boldsymbol{Y} \mid \boldsymbol{X}] \\
&= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{X}\beta \\
&= \beta.
\end{aligned}$$

- **In the linear regression model with i.i.d. sampling**

$$\mathbb{E}[\widehat{\beta} \mid \boldsymbol{X}] = \beta$$

- Using the law of iterative expectation, we can further prove that $\mathbb{E}[\widehat{\beta}] = \beta$.

Variance of Least Squares Estimator

君须记：(A^{-1})^T=(A^T)^{-1}

var[Z]ij=cov(Z_iZ_j)

- For any $r \times 1$ random vector Z, define the $r \times r$ covariance matrix

此处非转置

$$\operatorname{var}[Z] = \mathbb{E}\left[(Z - \mathbb{E}[Z])(Z - \mathbb{E}[Z])'\right] = \mathbb{E}\left[ZZ'\right] - (\mathbb{E}[Z])(\mathbb{E}[Z])'$$

- For any pair (Z, X), define the conditional covariance matrix

$$\operatorname{var}[Z \mid X] = \mathbb{E}\left[(Z - \mathbb{E}[Z \mid X])(Z - \mathbb{E}[Z \mid X])' \mid X\right]$$

- We define $\boldsymbol{V}_{\widehat{\beta}} \overset{\text{def}}{=} \operatorname{var}[\widehat{\beta} \mid \boldsymbol{X}]$ as the conditional covariance matrix of the regression coefficient estimators.
- We now derive its form.

详见照片

Variance of Least Squares Estimator

- The conditional covariance matrix of the $n \times 1$ regression error $\boldsymbol{e}$ is the $n \times n$ matrix.

$$\operatorname{var}[\boldsymbol{e} \mid \boldsymbol{X}] = \mathbb{E}\left[\boldsymbol{e}\boldsymbol{e}' \mid \boldsymbol{X}\right] \stackrel{\text{def}}{=} \boldsymbol{D}$$

var[e|x]_{ij}=E[e_ie_j|x]

- The i^{th} diagonal element of D is

E[e_i^2|X]=E[(e_i-E[e_i|x])^2|x]

$$\mathbb{E}\left[e_i^2 \mid \boldsymbol{X}\right] = \mathbb{E}\left[e_i^2 \mid X_i\right] = \sigma_i^2$$

- The ij^{th} off-diagonal element of D is

$$\mathbb{E}\left[e_i e_j \mid \boldsymbol{X}\right] = \mathbb{E}\left(e_i \mid X_i\right) \mathbb{E}\left[e_j \mid X_j\right] = 0$$

Variance of Least Squares Estimator

- Thus $\boldsymbol{D}$ is a diagonal matrix with i^{th} diagonal element σ_i^2 :

$$\boldsymbol{D} = \operatorname{diag}\left(\sigma_1^2, \ldots, \sigma_n^2\right) = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}$$

异方差+LRM

- In the special case of the linear homoskedastic regression model, then $\mathbb{E}\left[e_i^2 \mid X_i\right] = \sigma_i^2 = \sigma^2$ and we have the simplification $\boldsymbol{D} = \boldsymbol{I}_n\sigma^2$. In general, however, $\boldsymbol{D}$ need not necessarily take this simplified form.

同方差+LRM

Variance of Least Squares Estimator

- For any $n \times r$ matrix $A = A(X)$,

$$\operatorname{var}\left[\boldsymbol{A}'\boldsymbol{Y} \mid \boldsymbol{X}\right] = \operatorname{var}\left[\boldsymbol{A}'\boldsymbol{e} \mid \boldsymbol{X}\right] = \boldsymbol{A}'\boldsymbol{D}\boldsymbol{A}$$

- In particular, we can write $\widehat{\beta} = \boldsymbol{A}'\boldsymbol{Y}$ where $\boldsymbol{A} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$ and thus

$$\boldsymbol{V}_{\widehat{\beta}} = \operatorname{var}[\widehat{\beta} \mid \boldsymbol{X}] = \boldsymbol{A}'\boldsymbol{D}\boldsymbol{A} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$$

- It is useful to note that $\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X} = \sum_{i=1}^{n} X_i X_i' \sigma_i^2$, a weighted version of $\boldsymbol{X}'\boldsymbol{X}$.
- In the special case of the linear homoskedastic regression model, $\boldsymbol{D} = \boldsymbol{I}_n \sigma^2$, so $\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X} = \boldsymbol{X}'\boldsymbol{X}\sigma^2$, and the covariance matrix simplifies to $\boldsymbol{V}_{\widehat{\beta}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\sigma^2$

将"照片"结果化简得

```
V_{\beta}=(X'X)^{-1}X'DX(X'X)^{-1}
```

Gauss-Markov Theorem

- Write the homoskedastic linear regression model in vector format as

BLUE=Best+Linear+Unbiased+Estimator

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{e}$$
$$\mathbb{E}[\boldsymbol{e} \mid \boldsymbol{X}] = 0$$
$$\text{var}[\boldsymbol{e} \mid \boldsymbol{X}] = \boldsymbol{I}_n\sigma^2$$

- In this model we know that the least squares estimator is unbiased for β and has covariance matrix $\sigma^2(X'X)^{-1}$.
- **Is there an alternative unbiased estimator $\widetilde{\beta}$ which has a smaller covariance matrix?**

Linear:\Theta=A'Y,\theta=\sum a_i·y_i,
\beta=(X'X)^{-1}X'Y,A'=(X'X)^{-1}X'

Unbiased: if linear, it is unbiased

Gauss-Markov Theorem

- **Take the homoskedastic linear regression model. If $\widetilde{\beta}$ is an unbiased estimator of β then** $\text{var}[\widetilde{\beta} \mid \boldsymbol{X}] \geq \sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}$
- Since the variance of the OLS estimator is exactly equal to this bound this means that no unbiased estimator has a lower variance than OLS. Consequently we describe OLS as **efficient** in the class of unbiased estimators.
- Let's restrict attention to linear estimator of β, which are estimators that can be written as $\widetilde{\beta} = \boldsymbol{A}'\boldsymbol{Y}$, where $\boldsymbol{A} = \boldsymbol{A}(\boldsymbol{X})$ is an $m \times n$ function of the regressors $\boldsymbol{X}$.
- This restriction give rise to the description of OLS as the **best linear unbiased estimator (BLUE)**.

在同方差线性回归模型下，OLS估计量在所有可能的无偏估计量方差最小
弱证明：线性的
线性无偏估计量：\beta=A' · Y
即证:E(β |X)= β

Gauss-Markov Theorem

- For $\widetilde{\beta} = \boldsymbol{A}'\boldsymbol{Y}$ we have

$$\mathbb{E}[\widetilde{\beta} \mid \boldsymbol{X}] = \boldsymbol{A}'\mathbb{E}[\boldsymbol{Y} \mid \boldsymbol{X}] = \boldsymbol{A}'\boldsymbol{X}\beta$$

- Then $\widetilde{\beta}$ is unbiased for all β if (and only if) $\boldsymbol{A}'\boldsymbol{X} = \boldsymbol{I}_k$. Furthermore

$$\operatorname{var}[\widetilde{\beta} \mid \boldsymbol{X}] = \operatorname{var}\left[\boldsymbol{A}'\boldsymbol{Y} \mid \boldsymbol{X}\right] = \boldsymbol{A}'\boldsymbol{D}\boldsymbol{A} = \boldsymbol{A}'\boldsymbol{A}\sigma^2$$

E[(β-E(β|X))(β-E(β|
X))'|X]=E[(\hat{β}-β)(\hat{β}-β)'|X]=E[A'ee'A|
X]=

- the last equality using the homoskedasticity assumption. To establish the Theorem we need to show that for any such matrix A

$$\boldsymbol{A}'\boldsymbol{A} \geq \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$$

- Set $\boldsymbol{C} = \boldsymbol{A} - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$. Note that $\boldsymbol{X}'\boldsymbol{C} = 0$. We calculate that

$$\begin{aligned}
\boldsymbol{A}'\boldsymbol{A} - \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} &= \left(\boldsymbol{C} + \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right)'\left(\boldsymbol{C} + \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right) - \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \\
&= \boldsymbol{C}'\boldsymbol{C} + \boldsymbol{C}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} + \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{C} \\
&+ \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} - \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \\
&= \boldsymbol{C}'\boldsymbol{C} \geq 0.
\end{aligned}$$

不考

Generalized Least Squares

前述证明使用同方差假设，说明异方差情形下OLS不一定最佳

- Take the linear regression model in matrix format $\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{e}$
- Consider a generalized situation where the observation errors are possibly correlated and/or heteroskedastic.

$$\mathbb{E}[\boldsymbol{e} \mid \boldsymbol{X}] = 0$$
$$\operatorname{var}[\boldsymbol{e} \mid \boldsymbol{X}] = \Sigma\sigma^2$$

对角矩阵，对角线部分不相等

  for some $n \times n$ matrix $\Sigma > 0$ (possibly a function of $\boldsymbol{X}$) and some scalar σ^2. This includes the independent sampling framework where Σ is diagonal but allows for non-diagonal covariance matrices as well.
- Under these assumptions, we can calculate the expectation and variance of the OLS estimator:

$$\mathbb{E}[\widehat{\beta} \mid \boldsymbol{X}] = \beta$$
$$\operatorname{var}[\widehat{\beta} \mid \boldsymbol{X}] = \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \left(\boldsymbol{X}'\Sigma\boldsymbol{X}\right) \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$$

Generalized Least Squares

- In this case, the OLS estimator is not efficient. Instead, we develop the **Generalized Least Squares (GLS)** estimator of β.
- **When Σ is known**, take the linear model and pre-multiply by $\Sigma^{-1/2}$. This produces the equation $\widetilde{\boldsymbol{Y}} = \widetilde{\boldsymbol{X}}\beta + \widetilde{\boldsymbol{e}}$ where $\widetilde{\boldsymbol{Y}} = \Sigma^{-1/2}\boldsymbol{Y}$, $\widetilde{\boldsymbol{X}} = \Sigma^{-1/2}\boldsymbol{X}$, and $\widetilde{\boldsymbol{e}} = \Sigma^{-1/2}\boldsymbol{e}$.
- Consider OLS estimation of β in this equation.

$$
\begin{aligned}
\widetilde{\beta}_{\mathrm{gls}} &= \left(\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}}\right)^{-1}\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{Y}} \\
&= \left(\left(\Sigma^{-1/2}\boldsymbol{X}\right)'\left(\Sigma^{-1/2}\boldsymbol{X}\right)\right)^{-1}\left(\Sigma^{-1/2}\boldsymbol{X}\right)'\left(\Sigma^{-1/2}\boldsymbol{Y}\right) \\
&= \left(\boldsymbol{X}'\Sigma^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\Sigma^{-1}\boldsymbol{Y}
\end{aligned}
$$

E[\tidle{e}\tidle{e}'|X]
=E[\sigma^{-frac{1}{2}ee'\sigma^{-frac{1}{2}}]
=\sigma^{-frac{1}{2}}E[ee'|X]\sigma^{-frac{1}{2}
=I_n\sigma^2

- You can calculate that

$$
\mathbb{E}\left[\tilde{\beta}_{\mathrm{gls}} \mid \boldsymbol{X}\right] = \beta
$$

$$
\operatorname{var}\left[\widetilde{\beta}_{\mathrm{gls}} \mid \boldsymbol{X}\right] = \sigma^2\left(\boldsymbol{X}'\Sigma^{-1}\boldsymbol{X}\right)^{-1}
$$

Residuals

- What are some properties of the residuals $\widehat{e}_i = Y_i - X_i'\widehat{\beta}$ in the context of the linear regression model?
- Recall that $\widehat{\boldsymbol{e}} = \boldsymbol{M}\boldsymbol{e}$

$$\mathbb{E}[\widehat{\boldsymbol{e}} \mid \boldsymbol{X}] = \mathbb{E}[\boldsymbol{M}\boldsymbol{e} \mid \boldsymbol{X}] = \boldsymbol{M}\mathbb{E}[\boldsymbol{e} \mid \boldsymbol{X}] = 0$$

$$\operatorname{var}[\widehat{\boldsymbol{e}} \mid \boldsymbol{X}] = \operatorname{var}[\boldsymbol{M}\boldsymbol{e} \mid \boldsymbol{X}] = \boldsymbol{M}\operatorname{var}[\boldsymbol{e} \mid \boldsymbol{X}]\boldsymbol{M} = \boldsymbol{M}\boldsymbol{D}\boldsymbol{M}$$

- Under the assumption of conditional homoskedasticity

$$\operatorname{var}[\widehat{\boldsymbol{e}} \mid \boldsymbol{X}] = \boldsymbol{M}\sigma^2$$

单位矩阵

- In particular, for a single observation i we can find the variance of $\widehat{e}_i$ by taking the i^{th} diagonal element. Since the i^{th} diagonal element of $\boldsymbol{M}$ is $1 - h_{ii}$ we obtain

$$\operatorname{var}\left[\widehat{e}_i \mid \boldsymbol{X}\right] = \mathbb{E}\left[\hat{e}_i^2 \mid \boldsymbol{X}\right] = (1 - h_{ii})\,\sigma^2$$

- Can you show the conditional expectation and variance of the prediction errors $\widetilde{e}_i = Y_i - X_i'\widehat{\beta}_{(-i)}$?

投影矩阵的第i行第i列为：h_i i

Estimation of Error Variance

- The error variance $\sigma^2 = \mathbb{E}\left[e^2\right]$ can be a parameter of interest.
- One estimator is the sample average of the squared residuals:

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \widehat{e}_i^2$$

- We can calculate the expectation of $\widehat{\sigma}^2$:

$$\widehat{\sigma}^2 = \frac{1}{n}\boldsymbol{e}'\boldsymbol{M}\boldsymbol{e} = \frac{1}{n}\operatorname{tr}\left(\boldsymbol{e}'\boldsymbol{M}\boldsymbol{e}\right) = \frac{1}{n}\operatorname{tr}\left(\boldsymbol{M}\boldsymbol{e}\boldsymbol{e}'\right)$$

- Then

MM=M

$$\begin{aligned}
\mathbb{E}\left[\widehat{\sigma}^2 \mid \boldsymbol{X}\right] &= \frac{1}{n}\operatorname{tr}\left(\mathbb{E}\left[\boldsymbol{M}\boldsymbol{e}\boldsymbol{e}' \mid \boldsymbol{X}\right]\right) \\
&= \frac{1}{n}\operatorname{tr}\left(\boldsymbol{M}\mathbb{E}\left[\boldsymbol{e}\boldsymbol{e}' \mid \boldsymbol{X}\right]\right) \\
&= \frac{1}{n}\operatorname{tr}(\boldsymbol{M}\boldsymbol{D}) \\
&= \frac{1}{n}\sum^{n}\left(1 - h_{ii}\right)\sigma_i^2
\end{aligned}$$

Estimation of Error Variance

- Adding the assumption of conditional homoskedasticity

$$\mathbb{E}\left[\widehat{\sigma}^2 \mid \boldsymbol{X}\right] = \frac{1}{n}\operatorname{tr}\left(M\sigma^2\right) = \sigma^2\left(\frac{n-k}{n}\right)$$

M的迹n-k

which means that $\widehat{\sigma}^2$ is biased towards zero.

- So we can define an unbiased estimator by rescaling

$$s^2 = \frac{1}{n-k}\sum_{i=1}^{n}\widehat{e}_i^2$$

- By the above calculation $\mathbb{E}\left[s^2 \mid \boldsymbol{X}\right] = \sigma^2$ and $\mathbb{E}\left[s^2\right] = \sigma^2$. Hence the estimator s^2 is unbiased for σ^2. Consequently, s^2 is known as the bias-corrected estimator for σ^2 and in empirical practice s^2 is the most widely used estimator for σ^2.

Covariance Matrix Estimation Under Homoskedasticity

- For inference we need an estimator of the covariance matrix $\boldsymbol{V}_{\widehat{\beta}}$ of the least squares estimator.
- Under homoskedasticity the covariance matrix takes the simple form

$$\boldsymbol{V}_{\widehat{\beta}}^{0} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\sigma^2$$

- Replacing σ^2 with its estimator s^2, we have

$$\widehat{\boldsymbol{V}}_{\widehat{\beta}}^{0} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}s^2$$

- Since s^2 is conditionally unbiased for σ^2 it is simple to calculate that $\widehat{\boldsymbol{V}}_{\widehat{\beta}}^{0}$ is conditionally unbiased for $\boldsymbol{V}_{\widehat{\beta}}$ **under the assumption of homoskedasticity**:

$$\mathbb{E}\left[\widehat{\boldsymbol{V}}_{\widehat{\beta}}^{0} \mid \boldsymbol{X}\right] = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\mathbb{E}\left[s^2 \mid \boldsymbol{X}\right] = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\sigma^2 = \boldsymbol{V}_{\widehat{\beta}}$$

Covariance Matrix Estimation Under Homoskedasticity

- This was the dominant covariance matrix estimator in applied econometrics for many years and is still the default method in most regression packages.
- Stata uses the covariance matrix estimator by default in linear regression unless an alternative is specified.
- However, the above covariance matrix estimator can be highly biased if homoskedasticity fails.

Covariance Matrix Estimation Under Heteroskedasticity

- Recall that the general form for the covariance matrix is

$$\boldsymbol{V}_{\widehat{\beta}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X}\right)\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$$

- This depends on the unknown matrix $\boldsymbol{D}$: $\boldsymbol{D} = \operatorname{diag}\left(\sigma_1^2, \ldots, \sigma_n^2\right) = \mathbb{E}\left[\boldsymbol{e}\boldsymbol{e}' \mid \boldsymbol{X}\right]$
- An ideal but infeasible estimator is

$$\begin{aligned} \widehat{\boldsymbol{V}}_{\widehat{\beta}}^{\text{ideal}} &= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\widetilde{\boldsymbol{D}}\boldsymbol{X}\right)\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \\ &= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\sum_{i=1}^{n} X_i X_i' e_i^2\right)\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \end{aligned}$$

- You can verify that $\mathbb{E}\left[\widehat{\boldsymbol{V}}_{\widehat{\beta}}^{\text{ideal}} \mid \boldsymbol{X}\right] = \boldsymbol{V}_{\widehat{\beta}}$. However, the errors e_i^2 are unobserved.

Covariance Matrix Estimation Under Heteroskedasticity

- We can replace e_i^2 with the squared residuals $\widehat{e}_i^2$.

$$\widehat{\boldsymbol{V}}_{\widehat{\beta}}^{\mathrm{HC0}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \left(\sum_{i=1}^{n} X_i X_i' \widehat{e}_i^2\right) \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$$

- The label "HC" refers to "heteroskedasticity-consistent". The label "HC0" refers to this being the baseline heteroskedasticity-consistent covariance matrix estimator.
- We know, however, that $\widehat{e}_i^2$ is biased towards zero. Recall that to estimate the variance σ^2 the unbiased estimator s^2 scales the moment estimator $\hat{\sigma}^2$ by $n/(n-k)$. We make the same adjustment here:

$$\widehat{\boldsymbol{V}}_{\widehat{\beta}}^{\mathrm{HC1}} = \left(\frac{n}{n-k}\right) \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \left(\sum_{i=1}^{n} X_i X_i' \widehat{e}_i^2\right) \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$$

Standard Errors

$$s\left(\widehat{\beta}_j\right) = \sqrt{\widehat{\boldsymbol{V}}_{\widehat{\beta}_j}} = \sqrt{\left[\widehat{\boldsymbol{V}}_{\widehat{\beta}}\right]_{jj}}$$

speak_mino	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
exposure	.301288	.1884761	1.60	0.111	-.0701431	.6727192
gender	.065414	.0238468	2.74	0.007	.018419	.1124091
agri_hukou_3	.1387243	.0472013	2.94	0.004	.0457044	.2317443
_cons	.1212733	.10498	1.16	0.249	-.0856116	.3281583

Measures of Fit

- As we described in the previous chapter a commonly reported measure of regression fit is the regression R^2 defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \widehat{e}_i^2}{\sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2} = 1 - \frac{\widehat{\sigma}^2}{\widehat{\sigma}_Y^2}$$

样本方差

where $\hat{\sigma}_Y^2 = n^{-1} \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2$ .R^2 is an estimator of the population parameter

不是无偏估计量,n-1是，均值损失自由度
见照片

$$\rho^2 = \frac{\operatorname{var}\left[X'\beta\right]}{\operatorname{var}[Y]} = 1 - \frac{\sigma^2}{\sigma_Y^2}$$

- However, $\widehat{\sigma}^2$ and $\widehat{\sigma}_Y^2$ are biased. Theil (1961) proposed replacing these by the unbiased versions s^2 and $\widetilde{\sigma}_Y^2 = (n-1)^{-1} \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2$ yielding what is known as R-bar-squared or adjusted R-squared:

无限解释变量趋于1

$$\bar{R}^2 = 1 - \frac{s^2}{\widetilde{\sigma}_Y^2} = 1 - \frac{(n-1) \sum_{i=1}^{n} \widehat{e}_i^2}{(n-k) \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2}$$