ECON 480-3
LECTURE 2: MORE LINEAR REGRESSION

Ivan A. Canay
Northwestern University

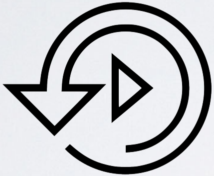
Past & Future

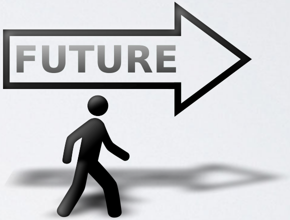
Last Class

- Interpretation of β in linear regression
- Solving and estimating β

Today

- Solving and estimating sub-vectors of β
- Properties of LS
- Estimating $\mathbb{V}$





Sub-Vectors of β

Let (Y, X, U) be a random vector where Y and U take values in $\mathbf{R}$ and X takes values in $\mathbf{R}^{k+1}$. Let $\beta = (\beta_0, \beta_1, \ldots, \beta_k)' \in \mathbf{R}^{k+1}$ be such that

$$Y = X'\beta + U \ .$$

Partition X into X_1 and X_2, where X_1 takes values in $\mathbf{R}^{k_1}$ and X_2 takes values in $\mathbf{R}^{k_2}$. Partition β into β_1 and β_2 analogously. In this notation,

$$Y = X_1'\beta_1 + X_2'\beta_2 + U \ .$$

Our preceding results imply that

$$\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} E[X_1X_1'] & E[X_1X_2'] \\ E[X_2X_1'] & E[X_2X_2'] \end{pmatrix}^{-1} \begin{pmatrix} E[X_1Y] \\ E[X_2Y] \end{pmatrix} \ .$$

Question: Can we derive formulae for β_1 and β_2 that admit some interesting interpretations?

Result Based on BLP

- **BLP**: for a random variable A and a random vector B, denote by $\text{BLP}(A|B)$ the best linear predictor of A given B, i.e.

$$\text{BLP}(A|B) \equiv B'E[BB']^{-1}E[BA] \; .$$

  If A is a random vector, then define $\text{BLP}(A|B)$ component-wise.

- Define $\tilde{Y} = Y - \text{BLP}(Y|X_2)$ and $\tilde{X}_1 = X_1 - \text{BLP}(X_1|X_2)$.

- Consider the linear regression

$$\tilde{Y} = \tilde{X}_1'\tilde{\beta}_1 + \tilde{U} \quad \text{where} \quad E[\tilde{X}_1\tilde{U}] = 0$$

  (as, for example, in the second interpretation of the linear regression model described before).

CLAIM

$$\tilde{\beta}_1 = E[\tilde{X}_1\tilde{X}_1']^{-1}E[\tilde{X}_1\tilde{Y}] = \beta_1$$

Result Based on BLP

Notation: BLP$(A|B) \equiv B'E[BB']^{-1}E[BA]$ and $\tilde{A} = A -$ BLP$(A|B)$.

Interpretation

- β_1 in the linear regression of Y on X_1 and X_2 is equal to the coefficient in a linear regression of the error term from a linear regression of Y on X_2 on the error terms from a linear regression of the components of X_1 on X_2.

- This formalizes the common description of β_1 as the "effect" of X_1 on Y after "controlling for X_2."

- Take $X_2 = constant$ and $X_1 \in \mathbf{R}$. Then $\tilde{Y} = Y - E[Y]$ and $\tilde{X}_1 = X_1 - E[X_1]$. Hence,

$$\begin{aligned} \beta_1 &= E[(X_1 - E[X_1])(X_1 - E[X_1])']^{-1} E[(X_1 - E[X_1])(Y - E[Y])] \\ &= \frac{\text{Cov}[X_1, Y]}{\text{Var}[X_1]} . \end{aligned}$$

- If we use our formula to interpret the coefficient β_j, we obtain

$$\beta_j = \frac{\text{Cov}[\tilde{X}_j, Y]}{\text{Var}[\tilde{X}_j]} . \tag{1}$$

$\Rightarrow$ each coefficient in a multivariate regression is the bivariate slope coefficient for the corresponding regressor, after "partialling out" all the other variables in the model.

QUESTIONS?

Estimating Sub-Vectors of β

- Partition X and β as before and consider

$$Y = X_1'\beta_1 + X_2'\beta_2 + U \ .$$

- Let $\mathbb{X}_1 = (X_{1,1}, \ldots, X_{1,n})'$ and $\mathbb{X}_2 = (X_{2,1}, \ldots, X_{2,n})'$.

- Denote by $\mathbb{P}_1$ the projection matrix onto the column space of $\mathbb{X}_1$ and $\mathbb{P}_2$ the projection matrix onto the column space of $\mathbb{X}_2$.

- Define $\mathbb{M}_1 = \mathbb{I} - \mathbb{P}_1$ and $\mathbb{M}_2 = \mathbb{I} - \mathbb{P}_2$.

- Denote by $\hat{\beta}_n = (\hat{\beta}_{1,n}', \hat{\beta}_{2,n}')'$ the LS estimator of β in a regression of Y on X.

- We now derive estimation counterparts to the previous results about solving for sub-vectors of β. That is, $\hat{\beta}_{1,n}$ can also be obtained from a “residualized” regression.

Frisch-Waugh-Lovell Decomposition

Start by noticing that: $\mathbb{Y} = \mathbb{X}_1 \hat{\beta}_{1,n} + \mathbb{X}_2 \hat{\beta}_{2,n} + \hat{\mathbb{U}}$ and recall that $\mathbb{M}_2 = \mathbb{I} - \mathbb{P}_2$.

QUESTIONS?

Properties of LS

- Let (Y, X, U) be a random vector where Y and U take values in $\mathbf{R}$ and X takes values in $\mathbf{R}^{k+1}$. Assume further that the first component of X is a constant equal to one. Let $\beta \in \mathbf{R}^{k+1}$ be such that

$$Y = X'\beta + U \ .$$

- Suppose that (1) $E[XU] = 0$, (2) $E[XX'] < \infty$, and that (3) there is no perfect collinearity in X.
- Denote by P the marginal distribution of (Y, X).
- Let $(Y_1, X_1), \ldots, (Y_n, X_n)$ be an i.i.d. sample of random vectors with distribution P.
- The properties we will discuss today are
  1. Bias
  2. Gauss-Markov Theorem
  3. Consistency
  4. Asymptotic Normality

Bias

Claim

Under the assumption (1') $E[U|X] = 0$ (i.e., $E[Y|X] = X'\beta$) it follows that $E[\hat{\beta}_n] = \beta$.

Gauss Markov Theorem

- Suppose (1') $E[U|X] = 0$ and that (4') $\text{Var}[U|X] = \sigma^2$.
- When $\text{Var}[U|X]$ is constant (and therefore does not depend on X) we say that U is *homoskedastic*. Otherwise, we say that U is *heteroskedastic*.
- **Guass-Markov Theorem:** under these assumptions the OLS estimator is "best" in the sense that it has the "smallest" value of $\text{Var}[\mathbb{A}'\mathbb{Y}|X_1, \ldots, X_n]$ among all estimators of the form

$$\mathbb{A}'\mathbb{Y}$$

for some matrix $\mathbb{A} = \mathbb{A}(X_1, \ldots, X_n)$ satisfying

$$E[\mathbb{A}'\mathbb{Y}|X_1, \ldots, X_n] = \beta \ .$$

- "smallest" is understood as the partial order obtained by $B \geqslant \tilde{B}$ iff $B - \tilde{B}$ is positive semi-definite.
- This class of estimators includes the OLS estimator as a special case (by setting $\mathbb{A}' = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}'$). The property is sometimes expressed as saying that OLS is the "best linear unbiased estimator (BLUE)" of β under these assumptions.

GMT: Reformulating what we need to show

The estimator is $\mathbb{A}'\mathbb{Y}$ for $\mathbb{A} = \mathbb{A}(X_1, \ldots, X_n)$ and satisfies $E[\mathbb{A}'\mathbb{Y}|X_1, \ldots, X_n] = \beta$.

GMT: Completing the argument

Want to show $\mathbb{A}'\mathbb{A} - (\mathbb{X}'\mathbb{X})^{-1}$ is positive semi-definite for any $\mathbb{A}$ satisfying $\mathbb{A}'\mathbb{X} = \mathbb{I}$.

Consistency

Under our three main assumptions, $\hat{\beta}_n \xrightarrow{P} \beta$ as $n \to \infty$

Asymptotic Normality

Suppose in addition to (1)-(3) that (4) $\mathrm{Var}[XU] = E[XX'U^2] < \infty$. Then, as $n \to \infty$,

$$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} N(0, \mathbb{V}) \quad \text{where} \quad \mathbb{V} = E[XX']^{-1} E[XX'U^2] E[XX']^{-1} .$$

QUESTIONS?

Estimation $\mathbb{V}$

- **Inference**: often requires a consistent estimator of
$$\mathbb{V} = E[XX']^{-1}E[XX'U^2]E[XX']^{-1} .$$

- Note that $\mathbb{V}$ has the so-called sandwich form.

- **“bread”:** can be consistently estimated as before.

- **“meat”:** consider the case where (1’) $E[U|X] = 0$ and (4’) $\text{Var}[U|X] = \sigma^2$ (i.e., homoskedasticity).
Under these conditions,
$$\text{Var}[XU] = E[XX'U^2] = E[XX']\sigma^2 .$$
Hence,
$$\mathbb{V} = E[XX']^{-1}\sigma^2 .$$
A natural choice of estimator is therefore
$$\hat{\mathbb{V}}_n = \left(\frac{1}{n}\sum_{1\leqslant i\leqslant n} X_i X_i'\right)^{-1} \hat{\sigma}_n^2 ,$$
where $\hat{\sigma}_n^2$ is a consistent estimator of σ^2.

Estimation $\mathbb{V}$ under homoskedasticity

A natural choice for an estimator of σ^2 is $\hat{\sigma}_n^2 = \frac{1}{n} \sum_{1 \leqslant i \leqslant n} \hat{U}_i^2$. **Claim**: $\hat{\sigma}_n^2 \xrightarrow{P} \sigma^2$.

Estimation $\mathbb{V}$ via HC estimator

- When we do not assume $\text{Var}[U|X] = \sigma^2$, a natural choice of estimator is

$$\hat{\mathbb{V}}_n = \left(\frac{1}{n} \sum_{1 \leqslant i \leqslant n} X_i X_i' \right)^{-1} \left(\frac{1}{n} \sum_{1 \leqslant i \leqslant n} X_i X_i' \hat{U}_i^2 \right) \left(\frac{1}{n} \sum_{1 \leqslant i \leqslant n} X_i X_i' \right)^{-1} . \tag{2}$$

- Later in the class we will prove that this estimator is consistent, i.e.,

$$\hat{\mathbb{V}}_n \xrightarrow{P} \mathbb{V} \text{ as } n \to \infty \,,$$

  regardless of the functional form of $\text{Var}[U|X]$. Only requires (1) to (4)

- This estimator is called the Heteroskedasticity Consistent (HC) estimator of $\mathbb{V}$.

- The standard errors used to construct t-statistics are the square roots of the diagonal elements of $\hat{\mathbb{V}}_n$, and this is the topic of the third part of this class.

- **Important:** by default, `Stata` and `R` report homoskedastic-only standard errors.

QUESTIONS?

Measures of Fit

Common Practice: report the measure of fit

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS} ,$$

where

$$\begin{aligned} TSS &= \sum_{1 \leqslant i \leqslant n} (Y_i - \bar{Y}_n)^2 \\ ESS &= \sum_{1 \leqslant i \leqslant n} (\hat{Y}_i - \bar{Y}_n)^2 \\ SSR &= \sum_{1 \leqslant i \leqslant n} \hat{U}_i^2 . \end{aligned}$$

- $R^2 = 1$ if and only if $SSR = 0$, i.e., $\hat{U}_i = 0$ for all $1 \leqslant i \leqslant n$.
- $R^2 = 0$ if and only if $ESS = 0$, i.e., $\hat{Y}_i = \bar{Y}_n$ for all $1 \leqslant i \leqslant n$.

- View $\frac{1}{n} \sum_{1 \leqslant i \leqslant n} (Y_i - \bar{Y}_n)^2$ as an estimator of
$$\text{Var}[Y_i] .$$
- View $\frac{1}{n} \sum_{1 \leqslant i \leqslant n} \hat{U}_i^2$ as an estimator of
$$\text{Var}[U_i] .$$
- R^2 may be then viewed as an estimator of
$$1 - \frac{\text{Var}[U_i]}{\text{Var}[Y_i]} .$$
Replacing these estimators with their unbiased counterparts yields "adjusted" R^2,
$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1} \frac{SSR}{TSS} .$$
- R^2 always increases with the inclusion of an additional regressor, whereas $\bar{R}^2$ may not.