EC508: Econometrics
Linear Regression with Multiple Regressors

Jean-Jacques Forneron

Spring, 2023

Boston University

Linear Regression with Multiple Regressors

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \cdots + \beta_k X_{k,i} + u_i, \quad i = 1, \ldots, n$$

- Predict Y_i using multiple variables $X_{1,i}, \ldots, X_{k,i}$, $k + 1 < n$.
- In matrix notation the linear model becomes:

$$Y_i = X_i'\beta + u_i, \quad i = 1, \ldots, n$$

where:

$$X_i = \begin{pmatrix} 1 \\ X_{1,i} \\ \vdots \\ X_{k,i} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$$

- We can extend all the results from a single to multiple regressors using matrix algebra

Preliminaries: Matrix Inverse

- **Inverse Matrix:** Let A be an $n \times n$ non-singular matrix (i.e. the rows/columns are linearly independent) then A^{-1} exists and:

$$AA^{-1} = A^{-1}A = I_n = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

Preliminaries: Matrix Derivative

- **Matrix Derivative 1/2:** Let a, b be conformable vectors such that $a'b$ is well defined then:

$$\frac{\partial a'b}{\partial b} = \frac{\partial}{\partial b}(a'b) = a$$

- **Matrix Derivative 2/2:** Let A be a matrix and b be a vector that are conformable so that $b'Ab$ is well defined then:

$$\frac{\partial b'Ab}{\partial b} = \frac{\partial}{\partial b}(b'Ab) = 2Ab$$

Preliminaries: Matrix Notation

- Let Y, X, u be such that:

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, X = \begin{pmatrix} X_1' \\ \vdots \\ X_n' \end{pmatrix} = \begin{pmatrix} 1 & X_{1,1} & \dots & X_{k,1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{1,n} & \dots & X_{k,n} \end{pmatrix}, u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$$

- Then we can re-write:

$$Y_i = X_i'\beta + u_i, \quad i = 1, \dots, n$$

as

$$Y = X\beta + u$$

where

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$$

OLS estimator with multiple regressors

The OLS estimator minimizes the Sum of Squared Residuals:

$$\min_{b_0, b_1, \ldots, b_k} \sum_{i=1}^{n} (Y_i - [b_0 + b_1 X_{1,i} + \cdots + b_k X_{k,i}])^2$$

and can be written in matrix form as:

$$\min_{b=(b_0, b_1, \ldots, b_k)'} (Y - Xb)'(Y - Xb)$$

now note that

$$(Y - Xb)'(Y - Xb) = \underbrace{Y'Y}_{\text{does not depend on b}} - \underbrace{2Y'Xb}_{\text{linear in b}} + \underbrace{b'X'Xb}_{\text{quadratic in b}}$$

OLS estimator with multiple regressors

The OLS estimator minimizes:

$$\min_{b=(b_0,b_1,\dots,b_k)'} \underbrace{Y'Y}_{\text{does not depend on b}} - \underbrace{2Y'Xb}_{\text{linear in b}} + \underbrace{b'X'Xb}_{\text{quadratic in b}}$$

First order condition (take derivative wrt b)

$$0 - 2X'Y + 2X'X\hat{\beta} = 0$$

Assuming $X'X$ is invertible, we get:

$$\hat{\beta} = \left(X'X\right)^{-1} X'Y = \left(\sum_{i=1}^{n} X_i X_i'\right)^{-1} \sum_{i=1}^{n} X_i Y_i$$

The Least Squares Assumptions for Multiple Regression (SW 6.5)

$$Y_i = X_i'\beta + u_i, \quad i = 1, \ldots, n$$

1. The conditional distribution of u given the X's has mean zero, that is:
$$\mathbb{E}(u_i | X_{1,i} = x_1, \ldots, X_{k,i} = x_k) = 0.$$
2. $(X_{1,i}, \ldots, X_{k,i}, Y_i), i = 1, \ldots, n$, are i.i.d.
3. Large outliers are unlikely: $X_1, \ldots, X_k$, and Y have four moments:
$$\mathbb{E}\left(X_1^4\right) < \infty, \ldots, \mathbb{E}\left(X_k^4\right) < \infty, \mathbb{E}\left(Y^4\right) < \infty.$$
4. There is no perfect multicollinearity.

Assumption #1: the conditional mean of u given the included Xs is zero.

$$\mathbb{E}(u_i | X_{1,i} = x_1, \ldots, X_{k,i} = x_k) = 0.$$

- This has the same interpretation as in regression with a single regressor.
- Failure of this condition leads to omitted variable bias, specifically, if an omitted variable
  1. belongs in the equation (so is in u) and
  2. is correlated with an included X
- then this condition fails and there is OV bias.
- The best solution, if possible, is to include the omitted variable in the regression.
- A second, related solution is to include a variable that controls for the omitted variable (discussed in Ch. 7)

Assumptions #2, #3

- Assumption #2: $(X_{1,i}, \ldots, X_{k,i}, Y_i), i = 1, \ldots, n$, are i.i.d.
  This is satisfied automatically if the data are collected by simple random sampling.
- Assumption #3: large outliers are rare (finite fourth moments)
  This is the same assumption as we had before for a single regressor. As in the case of a single regressor, OLS can be sensitive to large outliers, so you need to check your data (scatterplots!) to make sure there are no crazy values (typos or coding errors).

Assumption #4: There is no perfect multicollinearity

- **Perfect multicollinearity** is when one of the regressors is an exact linear function of the other regressors.
- For $n \geq k+1$ this implies that X has full rank $(k+1)$ so that $X'X$ is invertible.
- We can have perfect multicollinearity when we include the same regressor twice or when the regressors are perfectly correlated...
- when the regressors are very correlated (e.g. $\rho_{X_1,X_2} \simeq \pm 1$; i.e. close to perfect multicollinearity) then the large sample properties derived thereafter are not a good approximation of the distribution of the estimates

Properties of OLS

3 properties:

1. **Unbiasedness**: $\mathbb{E}(\hat{\beta}) = \beta$
2. **Consistency**: $\hat{\beta} \xrightarrow{p} \beta$; i.e. $\hat{\beta}_j \xrightarrow{p} \beta_j$ for $j = 0, \dots, k$
3. **Asymptotic Normality**:

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \Sigma)$$

where

$$\Sigma = \Sigma_X^{-1} \Sigma_v \Sigma_X^{-1}$$
$$\Sigma_X = \mathbb{E}(X_i X_i')$$
$$\Sigma_v = \mathbb{E}(v_i v_i'), \quad v_i = X_i u_i$$

Additional properties of OLS

2 additional properties:

4. **Best Linear Unbiased Estimator**: under homoskedasticity, $var(u_i|X_i) = \sigma_u$ does not depend on X_i, then OLS is BLUE
5. **Efficiency**: under normality of u_i, as in the single regressor case

Proofs: 1. Unbiasedness

First, recall that:

$$Y = X\beta + u, \quad \hat{\beta} = \left(X'X\right)^{-1} X'Y$$

so that:

$$\hat{\beta} = \left(X'X\right)^{-1} X'Y = \beta + \left(X'X\right)^{-1} X'u$$

taking expectations:

$$\mathbb{E}(\hat{\beta}) = \beta + \mathbb{E}(\left(X'X\right)^{-1} X'\mathbb{E}(u|X))$$

iid + conditional mean zero implies the last term is 0.

Proofs: 2. Consistency

First, recall that:

$$\hat{\beta} = \beta + \left(\frac{1}{n} \sum_{i=1}^{n} X_i X_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} X_i u_i$$

By the WLLN

$$\frac{1}{n} \sum_{i=1}^{n} X_i X_i' \xrightarrow{p} \mathbb{E}\left(X_i X_i' \right), \quad \frac{1}{n} \sum_{i=1}^{n} X_i u_i \xrightarrow{p} \mathbb{E}\left(X_i u_i \right) = 0$$

Continuous Mapping Theorem + Slutsky:

$$\left(\frac{1}{n} \sum_{i=1}^{n} X_i X_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} X_i u_i \xrightarrow{p} \mathbb{E}\left(X_i X_i' \right)^{-1} \mathbb{E}\left(X_i u_i \right) = 0$$

Conclusion: $\hat{\beta} \xrightarrow{p} \beta$

Proofs: 3. Asymptotic Normality

First, recall that:

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{1}{n} \sum_{i=1}^{n} X_i X_i' \right)^{-1} \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} X_i u_i \right)$$

Central Limit Theorem

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} X_i u_i \right) = \sqrt{n} \bar{v}_n \xrightarrow{d} \mathcal{N}(0, \Sigma_v)$$

Continuous Mapping Theorem + Slutsky:

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \Sigma)$$