Lecture 11: Properties of OLS in multiple regression

Economics 326 — Introduction to Econometrics II

Vadim Marmer, UBC

Multiple regression and OLS

- Consider the multiple regression model with k regressors:

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \ldots + \beta_k X_{k,i} + U_i.$$

- Let $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ be the OLS estimators: if

$$\hat{U}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1,i} - \hat{\beta}_2 X_{2,i} - \ldots - \hat{\beta}_k X_{k,i},$$

then

$$\sum_{i=1}^{n} \hat{U}_i = \sum_{i=1}^{n} X_{1,i}\hat{U}_i = \ldots = \sum_{i=1}^{n} X_{k,i}\hat{U}_i = 0.$$

Multiple regression and OLS

- As in Lecture 9, we can write $\hat{\beta}_1$ as

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \tilde{X}_{1,i} Y_i}{\sum_{i=1}^{n} \tilde{X}_{1,i}^2}, \text{ where}$$

  - $\tilde{X}_{1,i}$ are the fitted OLS residuals: $\tilde{X}_{1,i} = X_{1,i} - \hat{\gamma}_0 - \hat{\gamma}_2 X_{2,i} - \ldots - \hat{\gamma}_k X_{k,i}$.
  - $\hat{\gamma}_0, \hat{\gamma}_2, \ldots, \hat{\gamma}_k$ are the OLS coefficients: $\sum_{i=1}^{n} \tilde{X}_{1,i} = \sum_{i=1}^{n} \tilde{X}_{1,i} X_{2,i} = \ldots = \sum_{i=1}^{n} \tilde{X}_{1,i} X_{k,i} = 0$.

- Similarly, we can write $\hat{\beta}_2$ as

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} \tilde{X}_{2,i} Y_i}{\sum_{i=1}^{n} \tilde{X}_{2,i}^2}, \text{ where}$$

  - $\tilde{X}_{2,i}$ are the fitted OLS residuals: $\tilde{X}_{2,i} = X_{2,i} - \hat{\delta}_0 - \hat{\delta}_1 X_{1,i} - \hat{\delta}_3 X_{3,i} - \ldots - \hat{\delta}_k X_{k,i}$.
  - $\hat{\delta}_0, \hat{\delta}_1, \hat{\delta}_3, \ldots, \hat{\delta}_k$ are the OLS coefficients: $\sum_{i=1}^{n} \tilde{X}_{2,i} = \sum_{i=1}^{n} \tilde{X}_{2,i} X_{1,i} = \sum_{i=1}^{n} \tilde{X}_{2,i} X_{3,i} = \ldots = \sum_{i=1}^{n} \tilde{X}_{2,i} X_{k,i} = 0$.

The OLS estimators are linear

- Consider $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n \tilde{X}_{1,i} Y_i}{\sum_{i=1}^n \tilde{X}_{1,i}^2} = \sum_{i=1}^n \frac{\tilde{X}_{1,i}}{\sum_{l=1}^n \tilde{X}_{1,l}^2} Y_i = \sum_{i=1}^n w_{1,i} Y_i,$$

  where

$$w_{1,i} = \frac{\tilde{X}_{1,i}}{\sum_{l=1}^n \tilde{X}_{1,l}^2}.$$

- Recall that $\tilde{X}_1$ are the residuals from a regression of X_1 on $X_2, \dots, X_k$ and a constant, and therefore $w_{1,i}$ depends only on $\mathbf{X}$.

Unbiasedness

- Suppose that
  1. $Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \dots + \beta_k X_{k,i} + U_i$.
  2. **Conditional on $\mathbf{X}$**, $\mathrm{E}\left[U_i \mid \mathbf{X}\right] = 0$ for all i's.
  - Conditioning on $\mathbf{X}$ means that we condition on all regressors for all observations: $\mathbf{X} = \{(X_{1,i}, X_{2,i}, \dots, X_{k,i}) : i = 1, \dots, n\}$.
- Under the above assumptions, conditional on $\mathbf{X}$:

$$\begin{aligned}
\mathrm{E}\left[\hat{\beta}_0 \mid \mathbf{X}\right] &= \beta_0, \\
\mathrm{E}\left[\hat{\beta}_1 \mid \mathbf{X}\right] &= \beta_1, \\
&\vdots \\
\mathrm{E}\left[\hat{\beta}_k \mid \mathbf{X}\right] &= \beta_k.
\end{aligned}$$

Proof of unbiasedness

- Substituting Y_i and expanding:

$$\begin{aligned}
\hat{\beta}_1 = \frac{\sum_{i=1}^n \tilde{X}_{1,i} Y_i}{\sum_{i=1}^n \tilde{X}_{1,i}^2} &= \frac{\sum_{i=1}^n \tilde{X}_{1,i} \left(\beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \dots + \beta_k X_{k,i} + U_i\right)}{\sum_{i=1}^n \tilde{X}_{1,i}^2} \\
&= \beta_0 \frac{\sum_{i=1}^n \tilde{X}_{1,i}}{\sum_{i=1}^n \tilde{X}_{1,i}^2} + \beta_1 \frac{\sum_{i=1}^n \tilde{X}_{1,i} X_{1,i}}{\sum_{i=1}^n \tilde{X}_{1,i}^2} + \beta_2 \frac{\sum_{i=1}^n \tilde{X}_{1,i} X_{2,i}}{\sum_{i=1}^n \tilde{X}_{1,i}^2} \\
&\quad + \dots + \beta_k \frac{\sum_{i=1}^n \tilde{X}_{1,i} X_{k,i}}{\sum_{i=1}^n \tilde{X}_{1,i}^2} + \frac{\sum_{i=1}^n \tilde{X}_{1,i} U_i}{\sum_{i=1}^n \tilde{X}_{1,i}^2}.
\end{aligned}$$

- Using the partitioned regression results from Lecture 9:

$$\begin{aligned}
&\sum_{i=1}^n \tilde{X}_{1,i} = \sum_{i=1}^n \tilde{X}_{1,i} X_{2,i} = \dots = \sum_{i=1}^n \tilde{X}_{1,i} X_{k,i} = 0, \\
&\sum_{i=1}^n \tilde{X}_{1,i} X_{1,i} = \sum_{i=1}^n \tilde{X}_{1,i}^2.
\end{aligned}$$

- Therefore,

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n \tilde{X}_{1,i} U_i}{\sum_{i=1}^n \tilde{X}_{1,i}^2}.$$

Proof of unbiasedness

- We have

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n \tilde{X}_{1,i} U_i}{\sum_{i=1}^n \tilde{X}_{1,i}^2}.$$

- **Conditional on X**,

$$\mathrm{E}\left[U_i \mid \mathbf{X}\right] = 0.$$

- Therefore, **conditional on X**,

$$\begin{aligned}
\mathrm{E}\left[\hat{\beta}_1 \mid \mathbf{X}\right] &= \mathrm{E}\left[\beta_1 + \frac{\sum_{i=1}^n \tilde{X}_{1,i} U_i}{\sum_{i=1}^n \tilde{X}_{1,i}^2} \mid \mathbf{X}\right] \\
&= \beta_1 + \frac{\sum_{i=1}^n \tilde{X}_{1,i} \mathrm{E}\left[U_i \mid \mathbf{X}\right]}{\sum_{i=1}^n \tilde{X}_{1,i}^2} \\
&= \beta_1.
\end{aligned}$$

Conditional variance of the OLS estimators

- Suppose that:
  1. $Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \ldots + \beta_k X_{k,i} + U_i$.
  2. **Conditional on X**, $\mathrm{E}\left[U_i \mid \mathbf{X}\right] = 0$ for all i's.
  3. **Conditional on X**, $\mathrm{E}\left[U_i^2 \mid \mathbf{X}\right] = \sigma^2$ for all i's.
  4. **Conditional on X**, $\mathrm{E}\left[U_i U_j \mid \mathbf{X}\right] = 0$ for all $i \neq j$.
- The conditional variance of $\hat{\beta}_1$ given $\mathbf{X}$ is

$$\mathrm{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right) = \frac{\sigma^2}{\sum_{i=1}^n \tilde{X}_{1,i}^2}.$$

- **Gauss-Markov Theorem:** Under Assumptions 1–4, the OLS estimators are **BLUE**.

Derivation of the conditional variance

- We have $\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n \tilde{X}_{1,i} U_i}{\sum_{i=1}^n \tilde{X}_{1,i}^2}$.

- Conditional on $\mathbf{X}$,

$$\begin{aligned}
\operatorname{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right) &= \mathrm{E}\left[\left(\hat{\beta}_1 - \mathrm{E}\left[\hat{\beta}_1 \mid \mathbf{X}\right]\right)^2 \mid \mathbf{X}\right] \\
&= \mathrm{E}\left[\left(\frac{\sum_{i=1}^n \tilde{X}_{1,i} U_i}{\sum_{i=1}^n \tilde{X}_{1,i}^2}\right)^2 \mid \mathbf{X}\right] \\
&= \frac{1}{\left(\sum_{i=1}^n \tilde{X}_{1,i}^2\right)^2} \mathrm{E}\left[\left(\sum_{i=1}^n \tilde{X}_{1,i} U_i\right)^2 \mid \mathbf{X}\right] \\
&= \frac{1}{\left(\sum_{i=1}^n \tilde{X}_{1,i}^2\right)^2} \left(\sum_{i=1}^n \tilde{X}_{1,i}^2 \sigma^2 + \sum_{i \neq j} \tilde{X}_{1,i} \tilde{X}_{1,j} \cdot 0\right) \\
&= \frac{\sigma^2 \sum_{i=1}^n \tilde{X}_{1,i}^2}{\left(\sum_{i=1}^n \tilde{X}_{1,i}^2\right)^2} = \frac{\sigma^2}{\sum_{i=1}^n \tilde{X}_{1,i}^2}.
\end{aligned}$$

Conditional covariance of the OLS estimators

- Consider $\hat{\beta}_1$ and $\hat{\beta}_2$:

$$\begin{aligned}
\hat{\beta}_1 &= \beta_1 + \frac{\sum_{i=1}^n \tilde{X}_{1,i} U_i}{\sum_{i=1}^n \tilde{X}_{1,i}^2}, \\
\hat{\beta}_2 &= \beta_2 + \frac{\sum_{i=1}^n \tilde{X}_{2,i} U_i}{\sum_{i=1}^n \tilde{X}_{2,i}^2},
\end{aligned}$$

  where
  - $\tilde{X}_1$ are the fitted residuals from the regression of X_1 on a constant and $X_2, X_3, \ldots, X_k$.
  - $\tilde{X}_2$ are the fitted residuals from the regression of X_2 on a constant and $X_1, X_3, \ldots, X_k$.
- We will show that given Assumptions 1–4, **conditional on X**:

$$\operatorname{Cov}\left(\hat{\beta}_1, \hat{\beta}_2 \mid \mathbf{X}\right) = \sigma^2 \frac{\sum_{i=1}^n \tilde{X}_{1,i} \tilde{X}_{2,i}}{\sum_{i=1}^n \tilde{X}_{1,i}^2 \sum_{i=1}^n \tilde{X}_{2,i}^2}$$

Conditional covariance of the OLS estimators

Conditional on $\mathbf{X}$,

$$\begin{aligned}
&\operatorname{Cov}\left(\hat{\beta}_1, \hat{\beta}_2 \mid \mathbf{X}\right) \\
&= \mathrm{E}\left[\left(\hat{\beta}_1 - \mathrm{E}\left[\hat{\beta}_1 \mid \mathbf{X}\right]\right)\left(\hat{\beta}_2 - \mathrm{E}\left[\hat{\beta}_2 \mid \mathbf{X}\right]\right) \mid \mathbf{X}\right] \\
&= \frac{1}{\sum_{i=1}^n \tilde{X}_{1,i}^2 \sum_{i=1}^n \tilde{X}_{2,i}^2} \mathrm{E}\left[\left(\sum_{i=1}^n \tilde{X}_{1,i} U_i\right)\left(\sum_{i=1}^n \tilde{X}_{2,i} U_i\right) \mid \mathbf{X}\right] \\
&= \frac{1}{\sum_{i=1}^n \tilde{X}_{1,i}^2 \sum_{i=1}^n \tilde{X}_{2,i}^2} \left(\sum_{i=1}^n \tilde{X}_{1,i} \tilde{X}_{2,i} \sigma^2 + \sum_{i \neq j} \tilde{X}_{1,i} \tilde{X}_{2,j} \cdot 0\right) \\
&= \sigma^2 \frac{\sum_{i=1}^n \tilde{X}_{1,i} \tilde{X}_{2,i}}{\sum_{i=1}^n \tilde{X}_{1,i}^2 \sum_{i=1}^n \tilde{X}_{2,i}^2}.
\end{aligned}$$

Normality of the OLS estimators

- In addition to Assumptions 1–4, assume that **conditional on $\mathbf{X}$**, U_i's are jointly normally distributed.
- $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ are linear estimators:

$$\hat{\beta}_j = \sum_{i=1}^{n} w_{j,i} Y_i = \beta_j + \sum_{i=1}^{n} w_{j,i} U_i,$$

  where

$$w_{j,i} = \frac{\tilde{X}_{j,i}}{\sum_{l=1}^{n} \tilde{X}_{j,l}^2},$$

  and $\tilde{X}_{j,i}$ are the residuals from the regression of X_j on the rest of the regressors.
- It follows that $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ are jointly normally distributed (conditional on $\mathbf{X}$).

Inclusion of irrelevant regressors

- Suppose that the true model is $Y_i = \beta_0 + \beta_1 X_{1,i} + U_i$.
- We could estimate β_1 by

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \left(X_{1,i} - \bar{X}_1\right) Y_i}{\sum_{i=1}^{n} \left(X_{1,i} - \bar{X}_1\right)^2}.$$

- Suppose that instead we regress Y on a constant, X_1, and $k-1$ additional regressors $X_2, \ldots, X_k$, i.e., we estimate β_1 by

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n} \tilde{X}_{1,i} Y_i}{\sum_{i=1}^{n} \tilde{X}_{1,i}^2}.$$

- We have

$$\begin{aligned}\tilde{\beta}_1 &= \frac{\sum_{i=1}^{n} \tilde{X}_{1,i} \left(\beta_0 + \beta_1 X_{1,i} + U_i\right)}{\sum_{i=1}^{n} \tilde{X}_{1,i}^2} \\ &= \beta_1 + \frac{\sum_{i=1}^{n} \tilde{X}_{1,i} U_i}{\sum_{i=1}^{n} \tilde{X}_{1,i}^2}.\end{aligned}$$

- Since conditional on $\mathbf{X}$, $\mathrm{E}\left[U_i \mid \mathbf{X}\right] = 0$, $\tilde{\beta}_1$ is **unbiased**!

Inclusion of irrelevant regressors

- When $Y_i = \beta_0 + \beta_1 X_{1,i} + U_i$,

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} \left(X_{1,i} - \bar{X}_1\right) Y_i}{\sum_{i=1}^{n} \left(X_{1,i} - \bar{X}_1\right)^2} \quad \text{and} \\
\tilde{\beta}_1 &= \frac{\sum_{i=1}^{n} \tilde{X}_{1,i} Y_i}{\sum_{i=1}^{n} \tilde{X}_{1,i}^2}
\end{aligned}
$$

are both unbiased.

- Conditional on $\mathbf{X}$,

$$
\begin{aligned}
\operatorname{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right) &= \frac{\sigma^2}{\sum_{i=1}^{n} \left(X_{1,i} - \bar{X}_1\right)^2} \quad \text{and} \\
\operatorname{Var}\left(\tilde{\beta}_1 \mid \mathbf{X}\right) &= \frac{\sigma^2}{\sum_{i=1}^{n} \tilde{X}_{1,i}^2}.
\end{aligned}
$$

- Since the true model has only X_1, by the Gauss-Markov Theorem $\hat{\beta}_1$ is BLUE and

$$
\operatorname{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right) \leq \operatorname{Var}\left(\tilde{\beta}_1 \mid \mathbf{X}\right).
$$

- Without the Gauss-Markov Theorem, one can show directly that $\sum_{i=1}^{n} \left(X_{1,i} - \bar{X}_1\right)^2 \geq \sum_{i=1}^{n} \tilde{X}_{1,i}^2$.

Proof of the variance inequality

- $\tilde{X}_{1,i}$ are the fitted residuals from regressing $X_{1,i}$ on a constant, $X_{2,i}, \ldots, X_{k,i}$:

$$
X_{1,i} = \hat{\gamma}_0 + \hat{\gamma}_2 X_{2,i} + \ldots + \hat{\gamma}_k X_{k,i} + \tilde{X}_{1,i}.
$$

- Consider the sums-of-squares for this regression:

$$
\begin{aligned}
SST_1 &= \sum_{i=1}^{n} \left(X_{1,i} - \bar{X}_1\right)^2, \\
SSE_1 &= \sum_{i=1}^{n} \left(\hat{\gamma}_0 + \hat{\gamma}_2 X_{2,i} + \ldots + \hat{\gamma}_k X_{k,i} - \bar{X}_1\right)^2, \\
SSR_1 &= \sum_{i=1}^{n} \tilde{X}_{1,i}^2.
\end{aligned}
$$

- Thus,

$$
\sum_{i=1}^{n} \left(X_{1,i} - \bar{X}_1\right)^2 - \sum_{i=1}^{n} \tilde{X}_{1,i}^2 = SST_1 - SSR_1 = SSE_1 \geq 0.
$$

Variance and the number of regressors

- In $Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1,i} + \hat{\beta}_2 X_{2,i} + ... + \hat{\beta}_k X_{k,i} + \hat{U}_i$, the variance of the OLS estimator $\hat{\beta}_1$ is

$$\text{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right) = \frac{\sigma^2}{\sum_{i=1}^{n} \tilde{X}_{1,i}^2} = \frac{\sigma^2}{SSR_1},$$

  where SSR_1 is the residual sum-of-squares from the regression of X_1 on a constant and the rest of the regressors.

- Since SSR_1 can only decrease when we add more regressors, $\text{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right)$ increases with k if the added regressors are irrelevant but correlated with the included regressors.
- If the added regressors are uncorrelated with X_1, inclusion of such regressors will not affect SSR_1 (in large samples) or the variance of $\hat{\beta}_1$.
- If the added regressors are uncorrelated with X_1 and affect Y, their inclusion will reduce σ^2 without affecting SSR_1 and will reduce the variance of $\hat{\beta}_1$.

Estimation of variances and covariances

- In $Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1,i} + \hat{\beta}_2 X_{2,i} + ... + \hat{\beta}_k X_{k,i} + \hat{U}_i$,

$$\begin{aligned}
\text{Var}\left(\hat{\beta}_1 \mid \mathbf{X}\right) &= \frac{\sigma^2}{\sum_{i=1}^{n} \tilde{X}_{1,i}^2}, \\
\text{Cov}\left(\hat{\beta}_1, \hat{\beta}_2 \mid \mathbf{X}\right) &= \sigma^2 \frac{\sum_{i=1}^{n} \tilde{X}_{1,i} \tilde{X}_{2,i}}{\sum_{i=1}^{n} \tilde{X}_{1,i}^2 \sum_{i=1}^{n} \tilde{X}_{2,i}^2}.
\end{aligned}$$

- Variances and covariances can be estimated by replacing σ^2 with

$$s^2 = \frac{1}{n-k-1} \sum_{i=1}^{n} \hat{U}_i^2.$$

- Estimated variance and covariance:

$$\begin{aligned}
\widehat{\text{Var}}\left(\hat{\beta}_1\right) &= \frac{s^2}{\sum_{i=1}^{n} \tilde{X}_{1,i}^2}, \\
\widehat{\text{Cov}}\left(\hat{\beta}_1, \hat{\beta}_2\right) &= s^2 \frac{\sum_{i=1}^{n} \tilde{X}_{1,i} \tilde{X}_{2,i}}{\sum_{i=1}^{n} \tilde{X}_{1,i}^2 \sum_{i=1}^{n} \tilde{X}_{2,i}^2}.
\end{aligned}$$