# Multiple Regression & Dummy Variables

## ECON0004: Introductory Econometrics & Microeconomics

UCL

University College London

March 2026

Based on: Wooldridge (2019) *Introductory Econometrics*, 7e; Lecture notes (4.1, 4.2, Extra Notes).

# Outline

1 Why Multiple Regression?

2 MLR Assumptions & BLUE

3 Interpretation & Functional Forms

4 Dummy Variables

5 Multicollinearity

6 Testing the Difference Between Coefficients

# The Limits of Simple Regression

Recall our running example from Lecture 1 (population model):

$$\ln w_i = \beta_0 + \beta_1 \, \text{edu}_i + u_i \qquad \text{(where } u_i \text{ contains ability and other factors)}$$

**Problem:** The simple regression **ignores ability** $A_i$.

- More-able workers earn higher wages *and* tend to get more education.
- If $A_i$ is omitted, $\hat{\beta}_1$ picks up the wage premium that is really due to ability — **omitted variable bias (OVB)**.

**Solution:** Add more regressors to **control for confounders** and recover a cleaner ceteris-paribus estimate of the education effect.

## Multiple Linear Regression (MLR) Model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + u_i$$

- $y_i$ — dependent variable (outcome)
- $x_{ji}$ — independent variable / regressor $j$ for unit $i$
- $\beta_j$ — population coefficient on $x_j$ (*unknown*)
- $u_i$ — error term capturing all other determinants

**Running example:**

$$\ln w_i = \beta_0 + \beta_1 \,\text{edu}_i + \beta_2 \,\text{exp}_i + u_i$$

Now $\beta_1$ measures the wage return to education *holding experience fixed*.

# OLS in Multiple Regression: SSR Minimisation

OLS chooses $(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)$ to minimise:

$$\text{SSR} = \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \cdots - \hat{\beta}_k x_{ki})^2$$

Taking $\partial \text{SSR} / \partial \hat{\beta}_j = 0$ for each $j = 0, 1, \ldots, k$ yields $k+1$ **normal equations**:

$$\sum_{i=1}^{n} \hat{u}_i = 0, \qquad \sum_{i=1}^{n} x_{ji}\, \hat{u}_i = 0 \quad (j = 1, \ldots, k)$$

The residuals are **orthogonal** to every regressor (and to the constant). These are the same properties as in simple regression — now in $k$ dimensions.

# MLR Assumptions: Overview

Five assumptions underpin OLS inference in multiple regression. Together they deliver **Best Linear Unbiased Estimator (BLUE)**.

| **Assumption** | **Name** | **What it rules out** |
|---|---|---|
| MLR.1 | Linearity in parameters | Non-linear parameters |
| MLR.2 | Random sampling | Selection/endogenous sampling |
| MLR.3 | No perfect collinearity | Redundant regressors |
| MLR.4 | Zero conditional mean | OVB, measurement error, simultaneity |
| MLR.5 | Homoskedasticity | Variance changing with $x$ |

# MLR.1 and MLR.2

**MLR.1 — Linearity in Parameters**

The population model is $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$, where $\beta_0, \ldots, \beta_k$ are unknown constants.

**MLR.2 — Random Sampling**

We observe a random sample $\{(x_{1i}, \ldots, x_{ki}, y_i) : i = 1, \ldots, n\}$ drawn from the population model.

**Note:** MLR.1 allows *non-linear* functions of $x$ (e.g. $x^2$, $\ln x$) as long as the *parameters* enter linearly.

# MLR.3 and MLR.4

**MLR.3 — No Perfect Collinearity**

In the sample, no regressor is an exact linear combination of the others, and each regressor has some variation ($\text{SST}_j > 0$).

**MLR.4 — Zero Conditional Mean**

$$\mathbb{E}(u_i \mid x_{1i}, \ldots, x_{ki}) = 0$$

MLR.4 is the **key identifying assumption**: all relevant determinants of $y$ are either included as regressors or are *uncorrelated* with every $x_j$. Common violations: (i) omitted variable bias, (ii) measurement error in regressors, (iii) simultaneous causality (reverse causality).

**MLR.5 — Homoskedasticity**

$\text{Var}(u_i \mid x_{1i}, \ldots, x_{ki}) = \sigma^2$ (constant)

**Gauss-Markov Theorem**

Under MLR.1–5, OLS is **BLUE**: among all *linear unbiased* estimators of $\beta_j$, OLS has the smallest variance. (Applies to cross-sectional data; time-series settings require an additional no-autocorrelation condition.)

**Variance formula:**

$$\text{Var}(\hat{\beta}_j \mid \mathbf{X}) = \frac{\sigma^2}{\text{SST}_j (1 - R_j^2)}$$

where $\text{SST}_j = \sum_i (x_{ji} - \bar{x}_j)^2$ and $R_j^2$ is the $R^2$ from regressing $x_j$ on all other regressors.

# Partialling Out: The FWL Theorem

**Question:** What exactly does $\hat{\beta}_1$ estimate in a multiple regression?

**Frisch–Waugh–Lovell (FWL) Theorem**

$\hat{\beta}_1$ from the multiple regression $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + u_i$ equals the slope from the *simple* regression of $\tilde{y}_i$ on $\tilde{x}_{1i}$, where:

- $\tilde{x}_{1i}$ = residuals from regressing $x_{1i}$ on all other regressors (including the constant)
- $\tilde{y}_i$ = residuals from regressing $y_i$ on all other regressors (including the constant)

OLS **partials out** the influence of all other regressors before estimating the effect of $x_1$. This is the formal meaning of "ceteris paribus."

# Partial Effects: Ceteris Paribus Interpretation

The fitted model:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \cdots$$

Suppose only $x_1$ changes by $\Delta x_1$, all others fixed:

$$\Delta \hat{y} = \hat{\beta}_1 \, \Delta x_1$$

- $\hat{\beta}_j$ = expected change in $y$ from a one-unit increase in $x_j$, **holding all other regressors constant**
- The intercept $\hat{\beta}_0$ does *not* affect the predicted change

**Example:** $\ln \ln w = \beta_0 + \beta_1 \, \text{edu} + \beta_2 \exp + u$, $\hat{\beta}_1$ is the return to one extra year of education among workers with the *same* level of experience.

# Log-Level Models: Semi-Elasticity

**Model:**

$$\ln y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

**Interpretation of $\hat{\beta}_1$:**

- Approximation (small changes): $\hat{\beta}_1 \times 100 \approx \%\Delta y$ for a 1-unit increase in $x_1$
- **Exact formula:** $\%\Delta y = 100[\exp(\hat{\beta}_1) - 1]$

**Example**

**Example:** Suppose $\hat{\beta}_1 = 0.08$ in a log-wage regression on years of education.

- Approximation: one extra year raises wages by $\approx 8\%$
- Exact: $100[\exp(0.08) - 1] \approx 100[1.0833 - 1] = 8.33\%$

# Log-Log Models: Elasticity

**Model:**

$$\ln y = \beta_0 + \beta_1 \ln x_1 + \beta_2 x_2 + u$$

**Interpretation:**

$$\hat{\beta}_1 = \frac{\%\Delta y}{\%\Delta x_1} \quad \Longrightarrow \quad \text{elasticity of } y \text{ w.r.t. } x_1$$

- A 1% increase in $x_1$ is associated with a $\hat{\beta}_1$% change in $y$, ceteris paribus
- Particularly useful for **price elasticity of demand (PED)**: $\hat{\beta}_1 < 0$ means demand falls when price rises (law of demand; $\hat{\beta}_1 > 0$ would indicate a Giffen good); $|\hat{\beta}_1| > 1$ implies elastic demand

**Quick guide to functional forms:** Level-level $\rightarrow$ units; log-level $\rightarrow$ %; level-log $\rightarrow$ units per %; log-log $\rightarrow$ elasticity.

# Quadratic Functional Form

**Model:**

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + u$$

The marginal effect of $x$ on $y$ *varies with* $x$:

$$\frac{dy}{dx} = \beta_1 + 2\beta_2 x$$

- $\beta_2 > 0$: increasing returns (effect grows larger)
- $\beta_2 < 0$: **diminishing returns** (effect shrinks); turning point at $x^* = -\dfrac{\beta_1}{2\beta_2}$

**Example**

**Example:** Wage-experience profile. $\ln w = \beta_0 + \beta_1 \exp + \beta_2 \exp^2 + u$ with $\hat{\beta}_1 > 0$ and $\hat{\beta}_2 < 0$.
Wages rise with experience but at a *decreasing* rate. Peak experience: $\exp^* = -\hat{\beta}_1/(2\hat{\beta}_2)$.

# Worked Example: Wage, Education, and Experience

Estimated log-wage equation (OLS, $n = 526$):

$$\widehat{\ln w} = 0.284 + 0.092\ \text{edu} + 0.041\ \text{exp} - 0.00071\ \text{exp}^2$$

**Interpretations:**

- One extra year of education raises wages by approximately 9.2%, ceteris paribus
- At 10 years of experience: marginal return is $0.041 - 2(0.00071)(10) = 0.041 - 0.0142 = 0.0268 \approx 2.68\%$ per year
- Peak experience: $\text{exp}^* = -\hat{\beta}_1/(2\hat{\beta}_2) = -(0.041)/[2 \times (-0.00071)] \approx 29$ years

Multiple regression lets us **hold experience constant** when estimating the education premium — something simple regression cannot do.

# What Is a Dummy Variable?

**Dummy (Indicator) Variable**

A binary variable $D_i \in \{0, 1\}$ that encodes a qualitative characteristic.

- $D_i = 1$ if the characteristic holds for unit $i$
- $D_i = 0$ otherwise (the **base/reference category**)

**Example:** Let $\text{female}_i = 1$ if worker $i$ is female, 0 if male.
Model:

$$\ln w_i = \beta_0 + \beta_1 \, \text{female}_i + \beta_2 \, \text{edu}_i + u_i$$

**Predicted log wages:**

$$\mathbb{E}[\ln w \mid \text{female} = 0, \text{edu}] = \beta_0 + \beta_2 \, \text{edu}$$

$$\mathbb{E}[\ln w \mid \text{female} = 1, \text{edu}] = (\beta_0 + \beta_1) + \beta_2 \, \text{edu}$$

# Interpreting the Dummy Coefficient

In $\ln w_i = \beta_0 + \beta_1 \text{female}_i + \beta_2 \text{edu}_i + u_i$:

- $\beta_1$ = the difference in (log) wages between females and males **with the same education level**
- The dummy *shifts the intercept* but leaves the slope $\beta_2$ unchanged
- If $\hat{\beta}_1 < 0$, females earn less on average conditional on education (raw gender gap, not necessarily discrimination)

**Why control for education?** If women and men differ systematically in their education levels, a simple comparison of average wages conflates the gender effect with education effects. The dummy in a multiple regression **isolates** the intercept shift.

# The Dummy Variable Trap

**Problem:** Suppose we include *both* $\text{female}_i$ and $\text{male}_i = 1 - \text{female}_i$.

Then $\text{female}_i + \text{male}_i = 1$ for every observation — an **exact linear combination** of the regressors and the constant. This violates **MLR.3** (no perfect collinearity) and OLS breaks down.

> **Rule:** For a categorical variable with $m$ categories, include $m - 1$ dummies. The omitted category becomes the **base group**; all coefficients are interpreted *relative to that base*.

Alternative: include all $m$ dummies but *drop the intercept*. Coefficients then equal the group means rather than differences from a base.

# Multiple Dummy Variables: Marital Status

**Four groups:** married male, married female, single male, single female.

Choose **single male** as the base category. Define:

- $\text{marriedmale}_i = 1$ if married male, 0 otherwise
- $\text{marriedfemale}_i = 1$ if married female, 0 otherwise
- $\text{singlefemale}_i = 1$ if single female, 0 otherwise

Model:

$$\ln w_i = \beta_0 + \beta_1 \,\text{marriedmale} + \beta_2 \,\text{marriedfemale} + \beta_3 \,\text{singlefemale} + \beta_4 \,\text{edu} + u_i$$

Each $\hat{\beta}_j$ measures the wage gap relative to single males, conditional on education.

# Interaction: Dummy $\times$ Continuous Variable

Allow the *slope* to differ by group:

$$\ln w_i = \beta_0 + \beta_1 \,\text{female}_i + \beta_2 \,\text{edu}_i + \beta_3 \,(\text{female}_i \times \text{edu}_i) + u_i$$

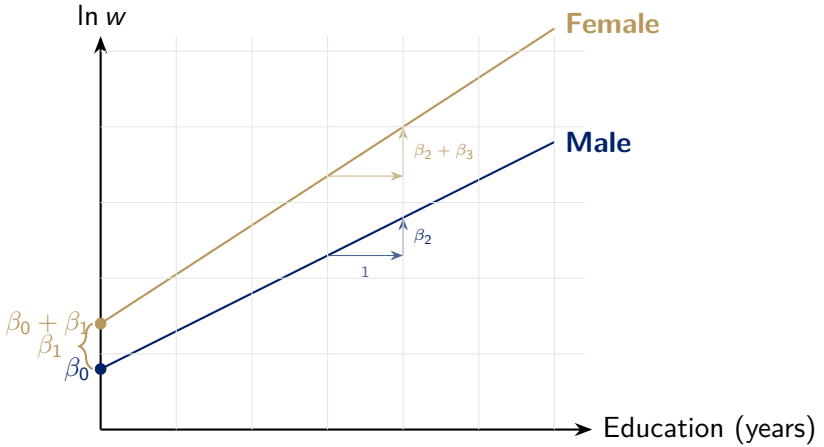
**Return to education by gender:**

$$\frac{\partial \ln w}{\partial \text{edu}} = \begin{cases} \beta_2 & \text{if male} \\ \beta_2 + \beta_3 & \text{if female} \end{cases}$$

| | **Male** | **Female** |
|---|---|---|
| Intercept | $\beta_0$ | $\beta_0 + \beta_1$ |
| Slope | $\beta_2$ | $\beta_2 + \beta_3$ |

$\beta_3 > 0$ would mean females have a *steeper* return to education than males.

# Visual: Interaction Model — Two Regression Lines




Both intercepts and slopes differ by gender. $\beta_1$ shifts the intercept; $\beta_3$ changes the return to education.

# Multicollinearity

**Imperfect multicollinearity:** regressors are highly (but not perfectly) correlated.

- OLS is still unbiased and consistent
- But the variance $\text{Var}(\hat{\beta}_j \mid \mathbf{X}) = \sigma^2/[\text{SST}_j(1 - R_j^2)]$ becomes **large** when $R_j^2 \to 1$ (regressor $j$ nearly explained by the others)
- Large SEs $\Rightarrow$ wide confidence intervals, $t$-statistics close to zero
- Coefficients may have unexpected signs or magnitudes

**MLR.3 — Perfect Collinearity (violated)**

If one regressor is an *exact* linear combination of the others, OLS cannot be computed: the normal equations have no unique solution.

**Example:** Including both "total household income" and the "sum of each member's income" — they are identical by construction.

# Testing $H_0 : \beta_1 = \beta_2$

**Example:** Does attending a 4-year university ($uni$ = years at university) give a higher return than a 2-year junior college ($jc$ = years at junior college)?
Model:

$$\ln w = \beta_0 + \beta_1 jc + \beta_2 uni + \beta_3 \exp + u$$

Test: $H_0 : \beta_1 = \beta_2$ vs $H_1 : \beta_1 < \beta_2$.

**Problem:** the standard $t$-test for $\hat{\beta}_1 - \hat{\beta}_2$ requires $\text{se}(\hat{\beta}_1 - \hat{\beta}_2)$, which needs $\text{Cov}(\hat{\beta}_1, \hat{\beta}_2)$ — not reported by default.

# The Reparameterisation Trick

Define $\theta_1 = \beta_1 - \beta_2$. Then $\beta_1 = \theta_1 + \beta_2$.
Substitute into the model:

$$\begin{aligned} \ln w &= \beta_0 + (\theta_1 + \beta_2)\, jc + \beta_2\, uni + \beta_3 \exp + u \\ &= \beta_0 + \theta_1\, jc + \beta_2\, (jc + uni) + \beta_3 \exp + u \end{aligned}$$

Regress $\ln w$ on $jc$, $(jc + uni)$, and exp. The coefficient on $jc$ is $\hat{\theta}_1$ and its standard error is reported directly.

> $H_0 : \beta_1 = \beta_2 \iff H_0 : \theta_1 = 0$.
>
> Run the reparameterised regression and apply a standard $t$-test on $\hat{\theta}_1$. No extra computation needed.

# Worked Example: 2-Year vs 4-Year College

Suppose we estimate the reparameterised model and find:

$$\hat{\theta}_1 = -0.031, \quad \text{se}(\hat{\theta}_1) = 0.012 \quad \Rightarrow \quad t = \frac{-0.031}{0.012} = -2.58$$

**Conclusion:**

- Critical value (one-sided, 5%, $df = n - 4$): $t^* \approx -1.645$
- Since $-2.58 < -1.645$, reject $H_0$ at the 5% level
- Junior college years have a significantly *lower* return than university years, ceteris paribus

> The reparameterisation trick converts a *linear* restriction ($\beta_1 = \beta_2$) into a single coefficient test, without requiring knowledge of $\text{Cov}(\hat{\beta}_1, \hat{\beta}_2)$.

# Lecture 4 Summary

1. **Multiple regression** controls for confounders and recovers ceteris-paribus estimates; OLS minimises SSR in $k+1$ dimensions.
2. **MLR.1–5** deliver BLUE. The key assumption is **zero conditional mean** (MLR.4); the variance formula shows collinearity inflates SEs.
3. **Functional forms:** log-level gives % effects; log-log gives elasticities; quadratic gives diminishing returns.
4. **Dummy variables** shift the intercept. The *dummy variable trap* requires omitting one category. Dummy $\times$ continuous interactions allow slope heterogeneity.
5. **Testing coefficient differences:** reparameterise to avoid needing $\text{Cov}(\hat{\beta}_1, \hat{\beta}_2)$.

**Next:** Even with multiple controls, omitted variables may still bias our estimates. Lecture 5 formalises **omitted variable bias** and introduces **quasi-experiments** (DiD) as a solution.