DegreeLive

Bachelor of Commerce (B2001) | Level 2 • Part B Major | Monash University

Introductory econometrics

Course Code: ETC2410 • Module 2 Notes

Module 2: Multiple Linear Regression, Inference & Dummy Variables

2.1 The Multiple Linear Regression Model and Partial Effects

Economic phenomena rarely depend on a single explanatory factor. When studying Australian residential property prices, household earnings, or ASX-listed firm profitability, omitting relevant determinants from a simple regression both reduces explanatory power and attributes their influence to the included regressor. The **Multiple Linear Regression (MLR)** model incorporates k explanatory variables simultaneously to achieve genuine *ceteris paribus* analysis:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + u_i$$

Where:

- y_i is the dependent variable for observation $i = 1, \dots, n$.
- $x_{i1}, x_{i2}, \dots, x_{ik}$ are the k explanatory variables for observation i.
- β_0 is the intercept parameter, representing $E(y_i \mid x_{i1} = 0, \dots, x_{ik} = 0)$ when $E(u_i \mid \mathbf{x}_i) = 0$.
- β_j $(j = 1, \dots, k)$ is the **partial regression coefficient** on x_j, measuring the expected change in y_i resulting from a one-unit increase in x_{ij} while holding all other included regressors x_{il} $(l \neq j)$ and the error term u_i constant: $\beta_j = \partial E(y_i \mid \mathbf{x}_i)/\partial x_{ij}$.
- u_i is the unobserved error term.

In matrix notation, stacking all n observations gives $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, where $\mathbf{X}$ is an $n \times (k+1)$ matrix (first column of ones followed by the k regressor columns) and $\boldsymbol{\beta} = [\beta_0, \beta_1, \dots, \beta_k]'$ is $(k+1) \times 1$. Minimising the Sum of Squared Residuals $SSR(\hat{\boldsymbol{\beta}}) = \hat{\mathbf{u}}'\hat{\mathbf{u}}$ yields the $(k+1)$ normal equations:

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y} \implies \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Where:

- $\mathbf{X}'\mathbf{X}$ is a $(k+1) \times (k+1)$ symmetric matrix, invertible whenever $\text{rank}(\mathbf{X}) = k+1$.
- $\hat{\boldsymbol{\beta}}$ is the $(k+1) \times 1$ vector of OLS parameter estimates.
- The normal equations imply $\mathbf{X}'\hat{\mathbf{u}} = \mathbf{0}$, meaning $\sum_{i=1}^{n} \hat{u}_i = 0$ and $\sum_{i=1}^{n} x_{ij}\hat{u}_i = 0$ for every regressor $j = 1, \dots, k$.

The Frisch-Waugh-Lovell (FWL) Partialling-Out Interpretation

To understand how OLS isolates the partial effect $\hat{\beta}_j$ when regressors are correlated with one another, consider the **Frisch-Waugh-Lovell Theorem**. For any single slope parameter $\hat{\beta}_1$ in a k-variable regression:

DegreeLive

1. Regress x_1 on all other explanatory variables $(x_2, \dots, x_k)$ and a constant, and obtain the residuals $\hat{r}_{i1}$. These residuals represent the component of x_{i1} that is linearly uncorrelated with $x_2, \dots, x_k$.
2. Regress y on $\hat{r}_1$ alone (with or without a constant, since $\bar{\hat{r}}_1 = 0$). The resulting simple regression slope equals $\hat{\beta}_1$ identically:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \hat{r}_{i1} y_i}{\sum_{i=1}^{n} \hat{r}_{i1}^2} = \frac{\sum_{i=1}^{n} \hat{r}_{i1} \tilde{y}_i}{\sum_{i=1}^{n} \hat{r}_{i1}^2}$$

Where:

- $\hat{r}_{i1}$ is the residual from regressing x_1 on $x_2, \dots, x_k$ and an intercept.
- $\tilde{y}_i$ is the residual from regressing y on $x_2, \dots, x_k$ and an intercept.

Thus, OLS automatically "partials out" the linear influence of all control variables from both x_1 and y before computing the bivariate slope. This formalises the *ceteris paribus* property in non-experimental economic data.

Worked Matrix Example: Australian Residential Property Valuation

To illustrate matrix OLS estimation in multiple regression, consider a small illustrative sample of $n = 5$ residential apartment sales in inner-city Melbourne, where y_i is sale price (AUD hundreds of thousands), x_{i1} is internal floor area (tens of square metres, e.g., 6 = 60 m²), and x_{i2} is distance from the Melbourne CBD (km).

Apartment (i)	Price (y_i, AUD 100k)	Floor Area (x_{i1}, 10 m²)	CBD Distance (x_{i2}, km)	OLS Fitted $\hat{y}_i$	OLS Residual $\hat{u}_i$
1	6.0	5	4	6.677	-0.677
2	8.0	7	3	7.903	0.097
3	7.5	7	6	6.871	0.629
4	9.5	9	2	9.129	0.371
5	9.0	12	5	9.419	-0.419

Step 1: Construct $\mathbf{X}'\mathbf{X}$ and $\mathbf{X}'\mathbf{y}$. With $n = 5$ and $k = 2$ slopes, $\mathbf{X}$ is 5×3 with a leading column of ones. Computing column sums and cross-products gives $\sum x_{i1} = 40$, $\sum x_{i2} = 20$, $\sum x_{i1}^2 = 348$, $\sum x_{i2}^2 = 90$, $\sum x_{i1}x_{i2} = 161$, $\sum y_i = 40.0$, $\sum x_{i1}y_i = 332.0$, and $\sum x_{i2}y_i = 157.0$:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 5 & 40 & 20 \\ 40 & 348 & 161 \\ 20 & 161 & 90 \end{bmatrix}, \qquad \mathbf{X}'\mathbf{y} = \begin{bmatrix} 40.0 \\ 332.0 \\ 157.0 \end{bmatrix}$$

Where $\mathbf{X}'\mathbf{X}$ is the 3×3 symmetric regressor cross-product matrix and $\mathbf{X}'\mathbf{y}$ is the 3×1 cross-product vector.

Step 2: Invert $\mathbf{X}'\mathbf{X}$ and solve $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Taking the matrix inverse (determinant $\det(\mathbf{X}'\mathbf{X}) = 1395$):

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{1395}\begin{bmatrix} 5399 & -380 & -520 \\ -380 & 50 & -5 \\ -520 & -5 & 140 \end{bmatrix} \approx \begin{bmatrix} 3.8703 & -0.2724 & -0.3728 \\ -0.2724 & 0.0358 & -0.0036 \\ -0.3728 & -0.0036 & 0.1004 \end{bmatrix}$$

Multiplying $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ yields $\hat{\beta}_0 = 5.8495$, $\hat{\beta}_1 = 0.4409$, and $\hat{\beta}_2 = -0.3441$. Holding CBD distance fixed, an extra 10 m² of floor area adds A$44,090 to apartment price; holding floor area fixed, each extra kilometre from the Melbourne CBD lowers price by A$34,410. The residual sum of squares is $SSR = \sum_{i=1}^{5} \hat{u}_i^2 = 1.1774$, yielding $\hat{\sigma}^2 = 1.1774/(5-2-1) = 0.5887$, $SE(\hat{\beta}_1) = \sqrt{0.5887 \times 0.03584} = 0.1453$, and $SE(\hat{\beta}_2) = \sqrt{0.5887 \times 0.10036} = 0.2431$.

2.2 MLR Assumptions, Adjusted R^2, and Estimator Variance

The Gauss-Markov framework extends directly from simple regression to multiple regression under assumptions **MLR.1 through MLR.5** (plus normality **MLR.6** for finite-sample exact inference):

- **MLR.1 (Linearity in Parameters):** $y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i$.
- **MLR.2 (Random Sampling):** $\{(x_{i1}, \ldots, x_{ik}, y_i)\}_{i=1}^{n}$ is an i.i.d. random sample.
- **MLR.3 (No Perfect Collinearity):** None of the regressors is constant, and no exact linear relationship exists among the explanatory variables, ensuring $\operatorname{rank}(\mathbf{X}) = k + 1$ and existence of $(\mathbf{X}'\mathbf{X})^{-1}$.
- **MLR.4 (Zero Conditional Mean):** $E(u_i \mid x_{i1}, \ldots, x_{ik}) = 0$.
- **MLR.5 (Homoskedasticity):** $\operatorname{Var}(u_i \mid x_{i1}, \ldots, x_{ik}) = \sigma^2$.
- **MLR.6 (Normality):** $u_i \mid \mathbf{x}_i \sim N(0, \sigma^2)$ independently across i.

Under MLR.1–MLR.4, $E(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) = \boldsymbol{\beta}$ (unbiasedness). Adding MLR.5 gives $\operatorname{Var}(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, and the Gauss-Markov Theorem establishes that $\hat{\boldsymbol{\beta}}$ is BLUE. The unbiased estimator of the error variance σ^2 adjusts for $k+1$ estimated parameters:

$$\hat{\sigma}^2 = \frac{SSR}{n - k - 1}$$

Where $n - k - 1$ is the residual degrees of freedom (n observations minus k slope coefficients and 1 intercept).

degreelive

Because adding any new explanatory variable to a regression can never increase SSR (and almost always reduces SSR slightly even if the variable is irrelevant), the unadjusted $R^2 = 1 - SSR/SST$ increases mechanically whenever regressors are added. To penalise unnecessary model complexity, econometricians report the **Adjusted** R^2 (denoted $\bar{R}^2$):

$$\bar{R}^2 = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)} = 1 - (1-R^2)\frac{n-1}{n-k-1}$$

Where:

- R^2 is the unadjusted coefficient of determination.
- $n-1$ is the total degrees of freedom and $n-k-1$ is the residual degrees of freedom.

Unlike R^2, $\bar{R}^2$ increases when a new regressor is added if and only if the $|t|$-statistic on that regressor exceeds 1 (meaning its inclusion reduces $\hat{\sigma}^2$). Note that $\bar{R}^2$ can occasionally be negative when R^2 is very close to zero and k is large relative to n.

2.3 Omitted Variable Bias (OVB)

Suppose the true population model contains two relevant variables, x_1 and x_2, satisfying MLR.1–MLR.4:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$$

Where β_1 and β_2 are the true partial effects. Suppose an analyst mistakenly omits x_2 (perhaps because it is unobservable, such as innate cognitive ability in a wage regression) and estimates the underspecified simple regression $\widetilde{y}_i = \widetilde{\beta}_0 + \widetilde{\beta}_1 x_{i1}$.

Using the algebraic formula for the simple OLS slope and substituting the true population equation for y_i, the conditional expectation of the underspecified estimator $\widetilde{\beta}_1$ is:

$$E(\widetilde{\beta}_1 \mid \mathbf{x}_1, \mathbf{x}_2) = \beta_1 + \beta_2 \widetilde{\delta}_1, \qquad \text{where } \widetilde{\delta}_1 = \frac{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)x_{i2}}{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2}$$

Where:

- $\widetilde{\beta}_1$ is the OLS slope from the short regression omitting x_2.
- β_1 is the true partial effect of x_1.
- β_2 is the true population effect of the omitted variable x_2 on y.
- $\widetilde{\delta}_1$ is the slope from an auxiliary regression of the omitted variable x_2 on the included variable x_1, capturing their sample correlation.

Degreelive

Therefore, the **Omitted Variable Bias** is $\text{Bias}(\tilde{\beta}_1) = \beta_2\tilde{\delta}_1$. Omitted variable bias vanishes ($extBias = 0$) only if at least one of two conditions holds: (1) x_2 has no partial effect on y ($\beta_2 = 0$), or (2) x_2 is completely uncorrelated with x_1 in the sample ($\tilde{\delta}_1 = 0$). Table 2.1 summarises the sign of the bias.

Sign of β_2 (Effect of x_2 on y)	$\text{Corr}(x_1, x_2) > 0$ ($\tilde{\delta}_1 > 0$)	$\text{Corr}(x_1, x_2) < 0$ ($\tilde{\delta}_1 < 0$)
$\beta_2 > 0$ (positive effect)	**Positive (upward) bias:** $E(\tilde{\beta}_1) > \beta_1$	**Negative (downward) bias:** $E(\tilde{\beta}_1) < \beta_1$
$\beta_2 < 0$ (negative effect)	**Negative (downward) bias:** $E(\tilde{\beta}_1) < \beta_1$	**Positive (upward) bias:** $E(\tilde{\beta}_1) > \beta_1$

Worked Example (Australian Returns to University Education): Consider estimating the return to schooling (*educ*, years of education) on $\ln(wage)$ using HILDA data while omitting unobserved individual ability (*abil*). Economic theory predicts that ability raises wages ($\beta_2 > 0$) and that higher-ability individuals complete more years of education ($\tilde{\delta}_1 > 0$). Suppose the true model is $\ln(wage) = \beta_0 + 0.065\,educ + 0.040\,abil + u$, and the auxiliary slope of *abil* on *educ* is $\tilde{\delta}_1 = 0.50$. Omitting *abil* yields an expected OLS schooling coefficient of 0.065 + (0.040)(0.50) = 0.085, overstating the true return to schooling by +0.020 (2.0 percentage points per year of education).

2.4 Multicollinearity and the Variance Inflation Factor (VIF)

While assumption MLR.3 prohibits *perfect* collinearity (for example, including both *income* in dollars and *income* in thousands of dollars, or including a full set of state dummies alongside an intercept), regressors in economic data are frequently strongly correlated with one another—a condition termed **imperfect multicollinearity**.

Under MLR.1–MLR.5, the conditional variance of any individual OLS slope estimator $\hat{\beta}_j$ ($j = 1, \dots, k$) can be written transparently using the FWL residual variation:

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

Where:

- $SST_j = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$ is the total sample variation of regressor x_j.
- R_j^2 is the R-squared obtained from an auxiliary regression of x_j on all other explanatory variables $x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_k$ and an intercept.
- $VIF_j = \frac{1}{1 - R_j^2}$ is the **Variance Inflation Factor** for regressor x_j.

Examining this variance formula clarifies three essential properties of multicollinearity:

1. **Unbiasedness and BLUE Remain Intact:** Imperfect multicollinearity ($R_j^2 < 1$) violates none of the Gauss-Markov assumptions MLR.1–MLR.5. OLS remains unbiased, consistent, and BLUE.
2. **Inflated Standard Errors and Low Power:** As R_j^2 approaches 1, VIF_j grows rapidly. For example, if $R_j^2 = 0.90$, $VIF_j = 10$, increasing the sampling variance tenfold and multiplying $SE(\hat{\beta}_j)$ by $\sqrt{10} \approx 3.16$. Consequently, individual t-statistics shrink ($t_j = \hat{\beta}_j / SE(\hat{\beta}_j)$), confidence intervals widen, and genuinely important predictors may appear individually insignificant even when the joint F-test is highly significant.
3. **Remedies and Pitfalls:** Dropping a relevant collinear variable solely to reduce VIF_j induces omitted variable bias, which is generally far worse than imprecision. Valid responses include expanding the sample size n (increasing SST_j), testing joint hypotheses via F-tests, or reframing collinear variables (for example, expressing household income and wealth per capita or combining strongly collinear items into an index).

2.5 Hypothesis Testing: t-Tests, Linear Combinations, and F-Tests

Under the Classical Linear Model assumptions MLR.1–MLR.6, the vector $\hat{\boldsymbol{\beta}}$ is multivariate normal conditional on $\mathbf{X}$, and individual standardised estimators follow exact Student t-distributions with $n - k - 1$ degrees of freedom.

Testing a Single Parameter and Confidence Intervals

To test the null hypothesis $H_0 : \beta_j = \beta_{j,0}$ against a two-sided alternative $H_1 : \beta_j \neq \beta_{j,0}$ (most commonly $\beta_{j,0} = 0$), the t-statistic is:

$$t = \frac{\hat{\beta}_j - \beta_{j,0}}{SE(\hat{\beta}_j)} \sim t_{n-k-1} \quad \text{under } H_0$$

Where $SE(\hat{\beta}_j) = \sqrt{\widehat{\text{Var}}(\hat{\beta}_j)}$ is the square root of the $(j+1)$-th diagonal element of $\hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$. We reject H_0 at significance level α if $|t| > t_{n-k-1,\alpha/2}$ (equivalently, if the two-sided p-value is less than α). The corresponding $(1-\alpha)$ confidence interval is $\hat{\beta}_j \pm t_{n-k-1,\alpha/2} \cdot SE(\hat{\beta}_j)$.

Testing a Single Linear Combination of Parameters

Economic hypotheses often compare two coefficients, such as testing whether a dollar of superannuation contribution affects household wealth equally to a dollar of non-super savings: $H_0 : \beta_1 = \beta_2$, rewritten as $H_0 : \theta = \beta_1 - \beta_2 = 0$. The point estimate is $\hat{\theta} = \hat{\beta}_1 - \hat{\beta}_2$, and its standard error requires the covariance between $\hat{\beta}_1$ and $\hat{\beta}_2$:

$$SE(\hat{\beta}_1 - \hat{\beta}_2) = \sqrt{\widehat{\text{Var}}(\hat{\beta}_1) + \widehat{\text{Var}}(\hat{\beta}_2) - 2\,\widehat{\text{Cov}}(\hat{\beta}_1, \hat{\beta}_2)}$$

DegreeLive

Where $\widehat{\text{Var}}(\hat{\beta}_1)$, $\widehat{\text{Var}}(\hat{\beta}_2)$, and $\widehat{\text{Cov}}(\hat{\beta}_1, \hat{\beta}_2)$ are extracted from $\hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$. Alternatively, reparameterising the model as $y_i = \beta_0 + \theta x_{i1} + \beta_2(x_{i1} + x_{i2}) + \cdots + u_i$ causes standard software to print $\hat{\theta}$ and $SE(\hat{\theta})$ directly.

The F-Test for Joint Linear Restrictions

Individual t-tests cannot test multiple parameter restrictions simultaneously (for instance, $H_0 : \beta_2 = 0$ and $\beta_3 = 0$), because doing separate t-tests inflates the family-wise Type I error rate and ignores correlation between $\hat{\beta}_2$ and $\hat{\beta}_3$. Instead, we compare the Residual Sum of Squares of an **unrestricted model** (SSR_{UR}, with k regressors and $df_{UR} = n - k - 1$) against a **restricted model** (SSR_R, estimated with the q linear restrictions of H_0 imposed):

$$F = \frac{(SSR_R - SSR_{UR})/q}{SSR_{UR}/(n-k-1)} = \frac{(R^2_{UR} - R^2_R)/q}{(1 - R^2_{UR})/(n-k-1)} \sim F_{q,\,n-k-1} \quad \text{under } H_0$$

Where:

- SSR_{UR} and R^2_{UR} are the residual sum of squares and R^2 from the full unrestricted regression.
- SSR_R and R^2_R are the residual sum of squares and R^2 from the restricted regression (only valid in the R^2 form when both models share the exact same dependent variable y).
- q is the number of independent linear restrictions under H_0 ($q = df_R - df_{UR}$).
- $n - k - 1$ is the denominator degrees of freedom from the unrestricted model.

Since imposing restrictions can never lower the sum of squared residuals ($SSR_R \geq SSR_{UR}$), the F-statistic is always non-negative, and rejection occurs strictly in the upper right tail whenever $F > F_{q,n-k-1,\alpha}$.

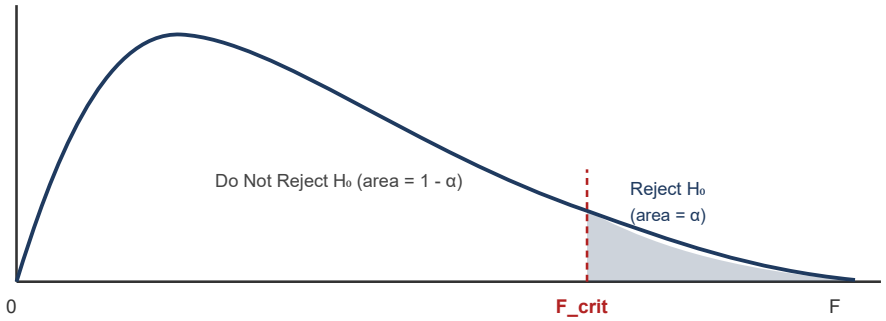


Figure 2.1: One-Sided Upper-Tail Rejection Region for the F_{q, n-k-1} Test

Worked Numerical F-Test (Australian ASX Firm Valuation): Suppose an analyst models the log market capitalisation ($\ln(mktcap)$) of $n = 126$ ASX-listed corporations as a function of log total assets ($\ln(assets)$), return on equity (roe), research and development intensity (rd_int), dividend payout ratio (div_pay), and leverage ($debt_ratio$). Thus $k = 5$ and $df_{UR} = 126 - 5 - 1 = 120$. Estimating the unrestricted model yields $SSR_{UR} = 36.00$ and $R^2_{UR} = 0.640$. To test the joint null

hypothesis that financial structure and payout variables have no incremental explanatory power (H_0 : $\beta_{rd_int} = 0$, $\beta_{div_pay} = 0$, $\beta_{debt_ratio} = 0$), we drop these $q = 3$ variables and estimate the restricted regression on $\ln(assets)$ and roe, obtaining $SSR_R = 40.50$ and $R_R^2 = 0.595$.

Calculation:

$$F = \frac{(40.50 - 36.00)/3}{36.00/120} = \frac{4.50/3}{0.30} = \frac{1.50}{0.30} = 5.00$$

Equivalently using the R^2 formula: $F = [(0.640 - 0.595)/3]/[(1 - 0.640)/120] = 0.015/0.003 = 5.00$. At the $\alpha = 1\%$ significance level, the critical value for $F_{3,120}$ is 3.95. Because $F = 5.00 > 3.95$, we reject H_0 at the 1% level and conclude that R&D intensity, dividend payout, and leverage are jointly statistically significant predictors of ASX firm market capitalisation.

A special case is the **overall regression F-statistic** reported in standard regression output, testing $H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$ ($q = k$, $R_R^2 = 0$), which simplifies to $F = [R^2/k]/[(1 - R^2)/(n - k - 1)]$.

2.6 Dummy Variables, Interactions, and Structural Change

Qualitative or categorical attributes—such as Australian state of residence, industry sector, sex, university degree attainment, or pre/post-COVID policy periods—enter regression models through binary **dummy (indicator) variables** equal to 1 if an observation belongs to a category and 0 otherwise.

Single Dummy Variable and Multi-Category Dummies

Consider a wage model incorporating a binary indicator $female_i$ (1 if female, 0 if male) alongside schooling ($educ$) and experience ($exper$):

$$\ln(wage_i) = \beta_0 + \delta_0 female_i + \beta_1 educ_i + \beta_2 exper_i + u_i$$

Where δ_0 measures the intercept shift between female and male workers holding education and experience constant (an approximate $100\delta_0\%$ differential, or exact $100[\exp(\delta_0) - 1]\%$ differential). Notice that we do *not* include both $female_i$ and $male_i$ when an intercept β_0 is present, because $female_i + male_i = 1$ creates exact collinearity with the constant column of ones—known as the **dummy variable trap**. In general, a categorical variable with m mutually exclusive categories requires exactly $m - 1$ dummy variables alongside the intercept, with the omitted category serving as the **reference (base) group**.

For example, comparing Australian full-time workers across New South Wales (reference group), Victoria (VIC_i), Queensland (QLD_i), and Western Australia (WA_i) requires three state dummies:

Degreelive

$$\widehat{\ln(wage)} = 2.40 - 0.03\,VIC + 0.05\,WA - 0.08\,QLD + 0.075\,educ$$

Where:

- The intercept 2.40 applies to the base state (NSW) at $educ = 0$.
- Conditional on schooling, Western Australian workers earn approximately 5% more than NSW workers (reflecting mining-sector wage premiums), whereas Queensland workers earn approximately 8% less than NSW workers.
- To compare WA against QLD directly, we test $H_0 : \delta_{WA} - \delta_{QLD} = 0$, yielding an estimated differential of 0.05 – (−0.08) = +0.13 (about 13% higher in WA than QLD).

Interaction Terms: Slope Shifts and Continuous Interactions

To allow the marginal effect of one variable to depend on another variable (either categorical or continuous), we include multiplicative **interaction terms**. First, interacting a dummy with a continuous regressor allows different slopes across groups:

$$\ln(wage_i) = \beta_0 + \delta_0 female_i + \beta_1 educ_i + \delta_1 (female_i \times educ_i) + u_i$$

Where:

- For males ($female_i = 0$): intercept is β_0 and return to education is β_1.
- For females ($female_i = 1$): intercept is $\beta_0 + \delta_0$ and return to education is $\beta_1 + \delta_1$.
- Testing whether the education slope differs by sex is a simple t-test of $H_0 : \delta_1 = 0$; testing whether the entire wage profile differs by sex (both intercept and slope) is a joint F-test of $H_0 : \delta_0 = 0,\ \delta_1 = 0$ ($q = 2$).

Second, interacting two continuous variables allows the partial effect of one regressor to vary linearly with the level of the other. In a Melbourne house price model ($price$ in AUD thousands, $size$ in square metres, $bedrooms$):

$$\widehat{price} = 210 + 3.20\,size + 28.0\,bedrooms + 0.25\,(size \times bedrooms)$$

Where the partial effect of an extra bedroom is $\partial \widehat{price}/\partial bedrooms = 28.0 + 0.25\,size$. Evaluated at the sample mean floor area $\overline{size} = 180$ m², an extra bedroom adds \$28.0 + 0.25(180) = 73.0($A 73,000). When reporting models with interaction terms, always evaluate partial effects at meaningful reference values (such as sample means) rather than interpreting main-effect coefficients at zero when zero is outside the data support.

DegreeLive

The Chow Test for Structural Stability

Suppose we wish to test whether all regression parameters in a model differ between two sub-samples—for instance, Australian retail sales sensitivity to interest rates before vs. after the 2020 pandemic (n_1 pre-2020 observations and n_2 post-2020 observations, with $n = n_1 + n_2$). Estimating separate OLS regressions on each sub-sample gives SSR_1 and SSR_2, so the unrestricted sum of squared residuals across both regimes is $SSR_{UR} = SSR_1 + SSR_2$ with $n - 2(k+1)$ degrees of freedom. Estimating the pooled regression without regime interactions yields SSR_P (the restricted SSR imposing $q = k+1$ equality constraints). The **Chow test statistic** is:

$$F = \frac{[SSR_P - (SSR_1 + SSR_2)]/(k+1)}{(SSR_1 + SSR_2)/[n - 2(k+1)]} \sim F_{k+1,\, n-2(k+1)} \quad \text{under } H_0$$

Where H_0 asserts that the intercept and all k slopes are identical across both periods. The Chow test is algebraically identical to an F-test on a regime dummy and all k dummy-regressor interactions in a single combined regression, assuming homoskedastic errors across both sub-samples.

2.7 Prediction: Confidence Intervals vs. Prediction Intervals

Beyond estimating marginal partial effects, economists frequently use fitted regressions to forecast an outcome at a specific target regressor profile $\mathbf{x}_0 = [1, x_{01}, \ldots, x_{0k}]'$. Crucially, the uncertainty surrounding the **conditional mean response** $E(y_0 \mid \mathbf{x}_0) = \mathbf{x}_0'\boldsymbol{\beta}$ differs from the uncertainty surrounding an **individual future realisation** $y_0 = \mathbf{x}_0'\boldsymbol{\beta} + u_0$.

The point prediction $\hat{y}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}}$ is identical in both cases, with sampling variance $\text{Var}(\hat{y}_0) = \sigma^2 \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0$. However, predicting an individual outcome incurs the additional variance of the unobserved future shock u_0, which is independent of the estimation sample:

$$SE(\hat{y}_0) = \hat{\sigma}\sqrt{\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}, \qquad SE(e_0) = \hat{\sigma}\sqrt{1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}$$

Where:

- $SE(\hat{y}_0)$ is the standard error for estimating the conditional mean $E(y_0 \mid \mathbf{x}_0)$, yielding the 95% confidence interval $\hat{y}_0 \pm t_{n-k-1,0.025} \cdot SE(\hat{y}_0)$.
- $SE(e_0)$ is the standard error of the prediction error $e_0 = y_0 - \hat{y}_0$, yielding the wider 95% prediction interval $\hat{y}_0 \pm t_{n-k-1,0.025} \cdot SE(e_0)$.

A convenient computational technique avoids manual matrix multiplication: re-centring every regressor around the target profile ($\tilde{x}_{ij} = x_{ij} - x_{0j}$) and regressing y_i on $\tilde{x}_{i1}, \ldots, \tilde{x}_{ik}$ produces an estimated intercept equal to $\hat{y}_0$ and an intercept standard error equal to $SE(\hat{y}_0)$ directly.

Worked Prediction Calculation (Melbourne Apartment): Using the Melbourne apartment regression from Section 2.1 ($n = 5$, $\hat{\sigma}^2 = 0.5887$, $df = 2$), predict the price of an 80 m² apartment ($x_{01} = 8$) located 4 km from the CBD ($x_{02} = 4$), matching sample means $\bar{x}_1 = 8$, $\bar{x}_2 = 4$. The point

forecast is $\hat{y}_0 = 5.8495 + 0.4409(8) - 0.3441(4) = 8.000$ (A$800,000). Because $\mathbf{x}_0$ lies exactly at the sample mean vector, $\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 = 1/n = 0.20$. Thus:

$$SE(\hat{y}_0) = \sqrt{0.5887 \times 0.20} = 0.3431, \qquad SE(e_0) = \sqrt{0.5887 \times (1 + 0.20)} = 0.8405$$

Using $t_{2,0.025} = 4.303$, the 95% confidence interval for the average price of all 80 m², 4 km apartments is 8.000 ± 4.303(0.3431) = [6.524, 9.476] (A$652,400 to A$947,600), whereas the 95% prediction interval for one specific apartment is [4.383, 11.617]. Notice how prediction uncertainty remains bounded below by $\hat{\sigma}$ even as $n \to \infty$.

2.8 Asymptotic Properties of OLS and the Lagrange Multiplier (LM) Test

In large Australian household or corporate datasets (such as HILDA or ABS Census microdata with thousands of observations), the finite-sample normality assumption MLR.6 is frequently violated—for example, wages are right-skewed and discrete indicators cannot be normal. Fortunately, large-sample (**asymptotic**) theory guarantees valid inference without MLR.6 under MLR.1–MLR.4 (or weaker zero-covariance exogeneity $E(u_i) = 0,\ \mathrm{Cov}(x_{ij}, u_i) = 0$):

- **Consistency:** By the Law of Large Numbers, $\operatorname{plim} \hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$ as $n \to \infty$, meaning the sampling distribution collapses around the true parameter vector.
- **Asymptotic Normality:** Under MLR.1–MLR.5, by the Central Limit Theorem, $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ converges in distribution to $N(\mathbf{0}, \sigma^2 \mathbf{Q}_{XX}^{-1})$, where $\mathbf{Q}_{XX} = \operatorname{plim}(\mathbf{X}'\mathbf{X}/n)$. Consequently, t-statistics converge to standard normal $N(0,1)$ distributions and $q \cdot F$ converges to χ^2_q.

An alternative large-sample test for q exclusion restrictions ($H_0 : \beta_{k-q+1} = \cdots = \beta_k = 0$) is the **Lagrange Multiplier (LM) test**, which estimates only the *restricted* model and checks whether its residuals correlate with the excluded variables:

1. Estimate the restricted regression of y on $x_1, \ldots, x_{k-q}$ and save residuals $\tilde{u}_i$.
2. Regress $\tilde{u}_i$ on *all* k regressors $(x_1, \ldots, x_k)$ and obtain the auxiliary $R^2_{\tilde{u}}$.
3. Compute $LM = n \cdot R^2_{\tilde{u}} \overset{a}{\sim} \chi^2_q$ under H_0, rejecting H_0 when $LM > \chi^2_{q,\alpha}$.

Although the F-test and LM test generally agree in large samples, the LM framework is indispensable in Chapter 3 for diagnostic testing of heteroskedasticity and serial correlation.

2.9 Policy Evaluation with Dummy Interactions: Difference-in-Differences

A premier application of categorical dummy variables and interaction terms in Australian applied economics is the **Difference-in-Differences (DiD)** estimator, used to evaluate state-level policy reforms (such as a Victorian payroll tax cut, a stamp duty concession, or minimum-wage adjustments) by comparing a treated state against a comparable control state before and after implementation.

Let $Treat_i = 1$ for units in the policy-treated group (e.g., Victoria) and 0 for the comparison group (e.g., New South Wales), and let $Post_t = 1$ in the post-policy period and 0 in the pre-policy period. Estimating the pooled regression:

$$y_{it} = \beta_0 + \beta_1 Treat_i + \beta_2 Post_t + \delta_{DiD}(Treat_i \times Post_t) + u_{it}$$

Where:

- β_0 is the baseline pre-policy mean for the control group ($Treat = 0, Post = 0$).
- β_1 captures permanent pre-existing differences between Victoria and NSW before the policy.
- β_2 captures common macroeconomic time shocks (such as RBA cash rate movements) affecting both states equally.
- δ_{DiD} is the Difference-in-Differences treatment effect, identifying the causal policy impact under the **parallel trends assumption** (that absent the policy, both states would have followed parallel trajectories).

Group	Before Policy ($Post = 0$)	After Policy ($Post = 1$)	Time Difference ($Post - Pre$)
Control (NSW, $Treat = 0$)	β_0	$\beta_0 + \beta_2$	β_2
Treated (VIC, $Treat = 1$)	$\beta_0 + \beta_1$	$\beta_0 + \beta_1 + \beta_2 + \delta_{DiD}$	$\beta_2 + \delta_{DiD}$
Between-Group Diff ($Treat - Control$)	β_1	$\beta_1 + \delta_{DiD}$	**DiD** $= \delta_{DiD}$

For example, if average regional SME quarterly hiring rises from 12.0 to 13.2 employees in NSW ($+1.2$) and from 11.5 to 14.1 employees in Victoria ($+2.6$) following a Victorian regional employment incentive, then $\hat{\delta}_{DiD} = (14.1 - 11.5) - (13.2 - 12.0) = 2.6 - 1.2 = +1.4$ additional employees per firm attributable to the policy, net of baseline state differences and national business-cycle shocks.

Degreelive