DegreeLive

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

Introductory econometrics

Course Code: ETC2410 • Module 1 Notes

Module 1: Simple Linear Regression Model, OLS Estimation & Classical Assumptions

1.1 The Scope of Econometrics and Economic Data Structures

Econometrics integrates economic theory, mathematical formulation, and statistical inference to quantify relationships among economic variables using observed empirical data. Unlike controlled laboratory experiments in the physical sciences, economic data are predominantly non-experimental (observational). Consequently, econometric methodology is structured to isolate causal or ceteris paribus associations—holding other relevant factors constant—while explicitly accounting for random disturbance and sampling variability.

In empirical economic research across Australia, such as policy analysis conducted by the Reserve Bank of Australia (RBA), the Treasury, or the Productivity Commission, datasets fall into four principal structural categories:

- **Cross-Sectional Data:** Observations collected across multiple individuals, households, firms, or regions at a single point in time (or within a short window where time ordering is ignored). An example is the Australian Bureau of Statistics (ABS) Survey of Income and Housing for a given financial year, recording weekly earnings, educational attainment, and state of residence across n sampled workers.
- **Time-Series Data:** Observations on one or more variables tracked sequentially over discrete time intervals (daily, monthly, quarterly, or annually). Examples include quarterly Australian real GDP growth, the RBA cash rate target, and monthly Consumer Price Index (CPI) inflation from 1990 to 2026. Time-series observations typically exhibit temporal dependence, violating simple random sampling assumptions.
- **Pooled Cross-Sectional Data:** Independent cross-sectional samples drawn from the same population across two or more distinct time periods. Combining the 2018 and 2024 ABS Household Expenditure Surveys forms a pooled cross-section that allows researchers to examine structural shifts in household savings behaviour.
- **Panel (Longitudinal) Data:** Repeated observations on the exact same cross-sectional units over multiple time periods. The Household, Income and Labour Dynamics in Australia (HILDA) Survey tracks the same Australians annually, enabling analysts to control for unobserved time-invariant individual characteristics such as innate ability or risk preferences.

Data Structure	Index Notation	Australian Empirical Example	Primary Econometric Challenge
Cross-Sectional	$i = 1, 2, \dots, n$	ABS Survey of Education and Work (2025)	Omitted variable bias; heteroskedasticity
Time-Series	$t = 1, 2, \dots, T$	RBA Cash Rate & CPI Quarterly Series	Serial correlation; non-stationarity; spurious trends
Pooled Cross-Section	$i = 1, \dots, n_t;\ t = 1, 2$	Combined 2019 & 2024 Census microdata samples	Changing population distributions over time

degreelive

Data Structure	Index Notation	Australian Empirical Example	Primary Econometric Challenge
Panel / Longitudinal	$i = 1, \dots, n;\ t = 1, \dots, T$	HILDA Survey annual labour force waves	Unobserved individual heterogeneity; attrition

1.2 Population Regression Function (PRF) vs. Sample Regression Function (SRF)

Consider a theoretical relationship between a dependent (explained or response) variable y and a single explanatory (independent or regressor) variable x. In economic analysis, y might represent annual household consumption expenditure in Australian dollars, while x represents annual household disposable income. Because identical income levels do not produce identical consumption across every household, the relationship is inherently stochastic rather than deterministic.

The **Simple Linear Regression Model** in the population is specified as:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

Where:

- y_i is the dependent variable for observation i.
- x_i is the independent (explanatory) variable for observation i.
- β_0 is the population intercept parameter, representing the expected value of y_i when $x_i = 0$ and the expected error is zero.
- β_1 is the population slope parameter, capturing the ceteris paribus change in y_i associated with a one-unit increase in x_i, holding the error term u_i fixed ($\Delta y_i / \Delta x_i = \beta_1$ when $\Delta u_i = 0$).
- u_i is the unobserved stochastic error (disturbance) term, absorbing all other factors affecting y_i besides x_i, such as household wealth, demographic composition, interest rate expectations, and measurement noise.

Assuming the conditional mean of the error term given x_i equals zero, $E(u_i \mid x_i) = 0$, taking the conditional expectation of both sides yields the **Population Regression Function (PRF)**:

$$E(y_i \mid x_i) = \beta_0 + \beta_1 x_i$$

Where:

- $E(y_i \mid x_i)$ is the conditional expectation of y_i given a specific value of x_i.
- $\beta_0 + \beta_1 x_i$ is the systematic (deterministic) component of y_i explained by x_i.

DegreeLive

The true population parameters β_0 and β_1 are unobservable because the analyst observes only a finite random sample $\{(x_i, y_i)\}_{i=1}^{n}$ drawn from the population. Using this sample, we construct estimators $\hat{\beta}_0$ and $\hat{\beta}_1$, which define the **Sample Regression Function (SRF)**:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

Where:

- $\hat{y}_i$ is the fitted (predicted) value of y_i for observation i.
- $\hat{\beta}_0$ is the sample OLS intercept estimate.
- $\hat{\beta}_1$ is the sample OLS slope estimate.

For any observed data point (x_i, y_i), the vertical deviation between the actual outcome y_i and the predicted value $\hat{y}_i$ is the **OLS residual**, denoted $\hat{u}_i$ (or e_i):

$$\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

Where:

- $\hat{u}_i$ is the observable sample residual for observation i, serving as the empirical counterpart to the unobservable population error u_i.
- y_i is the actual observed outcome and $\hat{y}_i$ is the fitted value on the sample regression line.

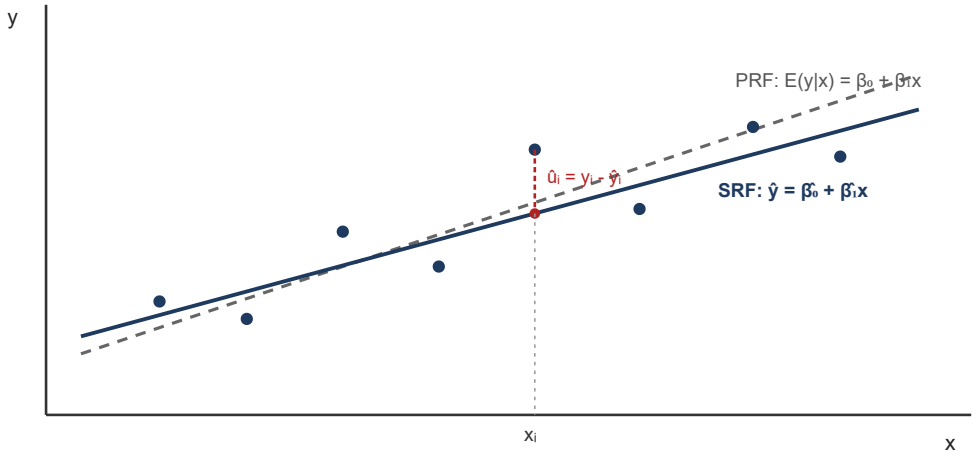


Figure 1.1: Population Regression Function (PRF), Sample Regression Function (SRF), and Residual $\hat{u}_i$

1.3 Ordinary Least Squares (OLS) Estimation and Normal Equations

The method of **Ordinary Least Squares (OLS)** selects the parameter estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ to minimise the Sum of Squared Residuals (SSR, also written RSS in some texts). Squaring residuals ensures that positive and negative deviations do not cancel out and penalises larger forecast errors more heavily than smaller ones:

$$\min_{\hat{\beta}_0, \hat{\beta}_1} SSR(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

Where:

- SSR is the Residual Sum of Squares (Sum of Squared Residuals) across all n sample observations.
- n is the sample size.
- $\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$ is the residual for the i-th observation.

To find the minimising values, we take partial derivatives of SSR with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$ and set them equal to zero. These first-order conditions yield the **OLS normal equations**:

$$\frac{\partial SSR}{\partial \hat{\beta}_0} = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \implies \sum_{i=1}^{n} \hat{u}_i = 0$$

$$\frac{\partial SSR}{\partial \hat{\beta}_1} = -2 \sum_{i=1}^{n} x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \implies \sum_{i=1}^{n} x_i \hat{u}_i = 0$$

Where:

- The first normal equation forces the sample residuals to sum to zero, $\sum_{i=1}^{n} \hat{u}_i = 0$, which implies their sample mean is zero ($\bar{\hat{u}} = 0$).
- The second normal equation forces the sample covariance between the regressor x_i and the residuals $\hat{u}_i$ to be zero ($\sum_{i=1}^{n} x_i \hat{u}_i = 0$).

Solving the first normal equation for $\hat{\beta}_0$ by dividing through by n gives the intercept estimator:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Where:

- $\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$ is the sample mean of the dependent variable.
- $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ is the sample mean of the explanatory variable.

Substituting $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ into the second normal equation and rearranging terms yields the closed-form OLS slope estimator, provided $\sum_{i=1}^{n} (x_i - \bar{x})^2 > 0$:

DegreeLive

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\underline{\text{Cov}}(x,y)}{\underline{\text{Var}}(x)}$$

Where:

- $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ is the sample cross-product of deviations of x and y from their respective means.
- $\sum_{i=1}^{n}(x_i - \bar{x})^2$ is the total sample variation in x around its mean.
- $\underline{\text{Cov}}(x,y)$ and $\underline{\text{Var}}(x)$ are the sample covariance and sample variance, respectively (with common $n-1$ denominators cancelling).

Algebraic Properties of OLS

Directly from the two normal equations and the definition of $\hat{y}_i$, five exact algebraic properties hold in every OLS regression containing an intercept, regardless of whether any probabilistic assumptions hold:

1. **Zero Mean Residual:** $\sum_{i=1}^{n}\hat{u}_i = 0$, and therefore $\bar{\hat{u}} = 0$.
2. **Zero Sample Covariance with x:** $\sum_{i=1}^{n} x_i\hat{u}_i = 0$, which implies $\sum_{i=1}^{n}(x_i - \bar{x})\hat{u}_i = 0$.
3. **Zero Sample Covariance with Fitted Values:** Because $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ is a linear function of x_i, $\sum_{i=1}^{n}\hat{y}_i\hat{u}_i = 0$.
4. **Point of Means on the Regression Line:** Substituting $x_i = \bar{x}$ into the SRF gives $\hat{y} = (\bar{y} - \hat{\beta}_1\bar{x}) + \hat{\beta}_1\bar{x} = \bar{y}$, so the point $(\bar{x},\bar{y})$ always lies on the OLS line.
5. **Mean of Fitted Values Equals Sample Mean:** Because $y_i = \hat{y}_i + \hat{u}_i$ and $\bar{\hat{u}} = 0$, it follows that $\bar{\hat{y}} = \bar{y}$.

1.4 Matrix Formulation of Ordinary Least Squares

Expressing the linear regression model in matrix algebra generalises seamlessly from simple regression to multiple regression and simplifies theoretical derivations of estimator properties. Stack the n observations into vectors and matrices:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

Where:

- $\mathbf{y}$ is the $n \times 1$ column vector of dependent variable observations $[y_1, y_2, \dots, y_n]'$.
- $\mathbf{X}$ is the $n \times 2$ design matrix whose first column consists of ones (for the intercept) and second column contains $[x_1, x_2, \dots, x_n]'$.
- $\boldsymbol{\beta} = [\beta_0, \beta_1]'$ is the 2×1 vector of unknown population parameters.
- $\mathbf{u} = [u_1, u_2, \dots, u_n]'$ is the $n \times 1$ vector of stochastic disturbances.

Degreelive

For any candidate parameter vector $\hat{\boldsymbol{\beta}}$, the residual vector is $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$, and the sum of squared residuals is the quadratic form:

$$SSR(\hat{\boldsymbol{\beta}}) = \hat{\mathbf{u}}'\hat{\mathbf{u}} = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{y}'\mathbf{y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}$$

Where:

- $\hat{\mathbf{u}}'$ denotes the transpose of the residual vector $\hat{\mathbf{u}}$.
- $\mathbf{X}'\mathbf{X}$ is the 2×2 symmetric cross-product matrix of regressors.
- $\mathbf{X}'\mathbf{y}$ is the 2×1 cross-product vector between regressors and the dependent variable.

Differentiating $SSR(\hat{\boldsymbol{\beta}})$ with respect to $\hat{\boldsymbol{\beta}}$ and setting the gradient to the zero vector gives the matrix normal equations:

$$\frac{\partial SSR}{\partial \hat{\boldsymbol{\beta}}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{0} \implies (\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$$

Where:

- $\mathbf{0}$ is the 2×1 zero vector.
- $(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$ is equivalent to $\mathbf{X}'\hat{\mathbf{u}} = \mathbf{0}$, directly encoding both summation normal equations.

Provided $\mathbf{X}$ has full column rank (meaning not all x_i are identical), $\mathbf{X}'\mathbf{X}$ is symmetric, positive definite, and invertible. Pre-multiplying by $(\mathbf{X}'\mathbf{X})^{-1}$ yields the matrix OLS estimator:

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

Where:

- $\hat{\boldsymbol{\beta}} = [\hat{\beta}_0, \hat{\beta}_1]'$ is the unique OLS estimator vector.
- $(\mathbf{X}'\mathbf{X})^{-1}$ is the inverse of the 2×2 cross-product matrix $\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}$.

1.5 Worked Numerical Example: Australian Household Consumption

Consider a sample of $n = 6$ Australian households surveyed in Melbourne, recording annual disposable income (x_i, in AUD thousands) and annual household consumption expenditure (y_i, in AUD thousands). We estimate the Keynesian consumption function $y_i = \beta_0 + \beta_1 x_i + u_i$, where β_1 represents the marginal propensity to consume (MPC).

DegreeLive

Household (i)	Income (x_i)	Consumption (y_i)	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
1	60	54	-40	-28	1120	1600
2	80	68	-20	-14	280	400
3	100	80	0	-2	0	0
4	100	84	0	2	0	0
5	120	96	20	14	280	400
6	140	110	40	28	1120	1600
Sum / Mean	$\sum x_i = 600$ ($\bar{x} = 100$)	$\sum y_i = 492$ ($\bar{y} = 82$)	0	0	**2800**	**4000**

Step 1: Compute sample means.

$$\bar{x} = \frac{600}{6} = 100 \text{ (AUD 100,000)}, \qquad \bar{y} = \frac{492}{6} = 82 \text{ (AUD 82,000)}$$

Where $\bar{x}$ and $\bar{y}$ are the sample arithmetic means of disposable income and consumption expenditure in thousands of Australian dollars.

Step 2: Compute the OLS slope $\hat{\beta}_1$ and intercept $\hat{\beta}_0$.

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{6}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{6}(x_i - \bar{x})^2} = \frac{2800}{4000} = 0.70$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 82 - (0.70)(100) = 12.0$$

Where $\hat{\beta}_1 = 0.70$ indicates that each additional A\$1,000 of annual household disposable income is associated with an estimated A\$700 increase in annual consumption expenditure (an MPC of 0.70). The intercept $\hat{\beta}_0 = 12.0$ corresponds to autonomous consumption of A\$12,000, though extrapolating to zero income lies far outside the observed range [60, 140].

Step 3: Verify via matrix notation. Using $\sum x_i = 600$, $\sum x_i^2 = 64000$, $\sum y_i = 492$, and $\sum x_i y_i = 52000$:

DegreeLive

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 6 & 600 \\ 600 & 64000 \end{bmatrix}, \qquad \mathbf{X}'\mathbf{y} = \begin{bmatrix} 492 \\ 52000 \end{bmatrix}$$

The determinant of $\mathbf{X}'\mathbf{X}$ is $(6)(64000) - (600)^2 = 384000 - 360000 = 24000$. Inverting the 2 × 2 matrix yields:

$$\hat{\boldsymbol{\beta}} = \frac{1}{24000}\begin{bmatrix} 64000 & -600 \\ -600 & 6 \end{bmatrix}\begin{bmatrix} 492 \\ 52000 \end{bmatrix} = \frac{1}{24000}\begin{bmatrix} 288000 \\ 16800 \end{bmatrix} = \begin{bmatrix} 12.0 \\ 0.70 \end{bmatrix}$$

Where the matrix solution matches the summation formulas exactly.

Step 4: Residuals and SSR. The fitted equation is $\hat{y}_i = 12 + 0.70x_i$. Evaluating at each observation gives $\hat{y} = [54, 68, 82, 82, 96, 110]'$ and residuals $\hat{u} = [0, 0, -2, 2, 0, 0]'$. Notice $\sum \hat{u}_i = 0$, $\sum x_i \hat{u}_i = 100(-2) + 100(2) = 0$, and $SSR = (-2)^2 + 2^2 = 8$.

1.6 Classical Linear Model (CLM) Assumptions & Gauss-Markov Theorem

While OLS estimates can always be computed mechanically whenever $\mathbf{X}'\mathbf{X}$ is invertible, statistical properties—such as unbiasedness, efficiency, and valid hypothesis tests—require probabilistic assumptions about the underlying population data-generating process. In the Simple Linear Regression (SLR) framework, the Gauss-Markov assumptions are:

Assumption	Formal Statement	Economic & Statistical Meaning
SLR.1 Linearity in Parameters	$y_i = \beta_0 + \beta_1 x_i + u_i$	The population model is linear in the unknown parameters β_0 and β_1 (variables may be nonlinearly transformed).
SLR.2 Random Sampling	$\{(x_i, y_i)\}_{i=1}^n$ is an i.i.d. random sample from the population.	Observations are mutually independent and identically distributed, typical of well-designed cross-sectional surveys.
SLR.3 Sample Variation in x	$\sum_{i=1}^n (x_i - \bar{x})^2 > 0$	The explanatory variable is not constant across the sample, ensuring a unique OLS slope exists.
SLR.4 Zero Conditional Mean	$E(u_i \mid x_i) = 0$	Unobserved factors in u_i have mean zero at every value of x_i; implies $\text{Cov}(x_i, u_i) = 0$ (exogeneity).
SLR.5 Homoskedasticity	$\text{Var}(u_i \mid x_i) = \sigma^2$	The conditional variance of the error term is constant across all levels of the explanatory variable x_i.

DegreeLive

Unbiasedness of OLS (SLR.1 – SLR.4)

An estimator $\hat{\theta}$ is **unbiased** for a population parameter θ if its expected value across repeated random samples equals the true parameter, $E(\hat{\theta}) = \theta$. Under assumptions SLR.1 through SLR.4, both $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased estimators of β_0 and β_1.

To prove unbiasedness of $\hat{\beta}_1$, let $s_x^2 = \sum_{i=1}^n (x_i - \bar{x})^2$ and rewrite $\hat{\beta}_1$ using $y_i = \beta_0 + \beta_1 x_i + u_i$ and $\sum(x_i - \bar{x}) = 0$:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{s_x^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i)}{s_x^2} = \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x}) u_i}{s_x^2}$$

Where:

- $s_x^2 = \sum_{i=1}^n (x_i - \bar{x})^2$ is the total sample variation in x.
- $\sum_{i=1}^n (x_i - \bar{x})\beta_0 = 0$ and $\sum_{i=1}^n (x_i - \bar{x}) x_i = \sum_{i=1}^n (x_i - \bar{x})^2$.

Conditioning on the sample values $\mathbf{x} = (x_1, \ldots, x_n)$ and applying SLR.4 ($E(u_i \mid \mathbf{x}) = 0$ by random sampling SLR.2):

$$E(\hat{\beta}_1 \mid \mathbf{x}) = \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x}) E(u_i \mid \mathbf{x})}{s_x^2} = \beta_1$$

Where $E(\hat{\beta}_1 \mid \mathbf{x}) = \beta_1$ implies by the Law of Iterated Expectations that the unconditional expectation satisfies $E(\hat{\beta}_1) = E[E(\hat{\beta}_1 \mid \mathbf{x})] = \beta_1$. Similarly, $E(\hat{\beta}_0 \mid \mathbf{x}) = \beta_0 + \beta_1 \bar{x} - \beta_1 \bar{x} = \beta_0$. If SLR.4 fails—for instance, because an omitted determinant of consumption such as household wealth is correlated with income—OLS becomes biased and inconsistent.

Sampling Variance of OLS Estimators (SLR.1 – SLR.5)

Adding the homoskedasticity assumption SLR.5 ($\mathrm{Var}(u_i \mid \mathbf{x}) = \sigma^2$) and noting that random sampling implies $\mathrm{Cov}(u_i, u_j \mid \mathbf{x}) = 0$ for $i \neq j$, the conditional sampling variances of the OLS slope and intercept are:

$$\mathrm{Var}(\hat{\beta}_1 \mid \mathbf{x}) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}, \qquad \mathrm{Var}(\hat{\beta}_0 \mid \mathbf{x}) = \frac{\sigma^2 \frac{1}{n} \sum_{i=1}^n x_i^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Where:

- $\sigma^2 = \mathrm{Var}(u_i \mid x_i)$ is the unknown constant variance of the population error term.
- $\sum_{i=1}^n (x_i - \bar{x})^2$ is the spread of x in the sample; larger variation in x shrinks estimator variance and improves precision.
- n is the sample size; larger n increases $\sum(x_i - \bar{x})^2$, reducing sampling variance at rate $1/n$.

DegreeLive

In matrix notation, the conditional variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ under homoskedasticity and uncorrelated errors is:

$$\text{Var}(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

Where:

- The diagonal entries of $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ contain $\text{Var}(\hat{\beta}_0 \mid \mathbf{X})$ and $\text{Var}(\hat{\beta}_1 \mid \mathbf{X})$.
- The off-diagonal entry contains $\text{Cov}(\hat{\beta}_0, \hat{\beta}_1 \mid \mathbf{X}) = -\sigma^2 \bar{x} / \sum (x_i - \bar{x})^2$.

Because σ^2 is unknown, it is estimated using the residual variance with a degrees-of-freedom adjustment for the two estimated parameters ($\hat{\beta}_0$ and $\hat{\beta}_1$):

$$\hat{\sigma}^2 = s^2 = \frac{SSR}{n-2} = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n-2}, \qquad SE(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

Where:

- $\hat{\sigma}^2$ (often denoted s^2) is an unbiased estimator of σ^2 under SLR.1–SLR.5, and its square root $\hat{\sigma}$ is the Standard Error of the Regression (SER).
- $n - 2$ represents the residual degrees of freedom, reflecting the two linear restrictions imposed by the OLS normal equations.
- $SE(\hat{\beta}_1)$ is the standard error of $\hat{\beta}_1$, measuring the estimated standard deviation of its sampling distribution.

The Gauss-Markov Theorem (BLUE)

Theorem (Gauss-Markov): Under assumptions SLR.1 through SLR.5, the Ordinary Least Squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are the **Best Linear Unbiased Estimators (BLUE)** of β_0 and β_1.

- **Linear:** Each estimator is a linear combination of the sample outcomes $y_1, \ldots, y_n$, written $\widetilde{\beta}_1 = \sum_{i=1}^{n} w_i y_i$ where weights w_i depend only on $\mathbf{x}$. For OLS, $w_i = (x_i - \bar{x}) / \sum_{j=1}^{n}(x_j - \bar{x})^2$.
- **Unbiased:** $E(\widetilde{\beta}_1 \mid \mathbf{x}) = \beta_1$ for all true parameter values, which requires $\sum_{i=1}^{n} w_i = 0$ and $\sum_{i=1}^{n} w_i x_i = 1$.
- **Best (Minimum Variance):** For any alternative linear unbiased estimator $\widetilde{\beta}_1$, $\text{Var}(\hat{\beta}_1) \leq \text{Var}(\widetilde{\beta}_1)$. In matrix form, $\text{Var}(\widetilde{\boldsymbol{\beta}}) - \text{Var}(\hat{\boldsymbol{\beta}})$ is positive semi-definite.

Exact Normality (SLR.6) and Classical Inference

Adding **SLR.6 (Normality)**—that the population error u_i is independent of x_i and normally distributed, $u_i \mid x_i \sim N(0, \sigma^2)$—completes the six Classical Linear Model (CLM) assumptions. Under SLR.1–SLR.6, $\hat{\beta}_1$ is normally distributed in finite samples, and replacing σ with $\hat{\sigma}$ yields an exact Student t-distribution:

DegreeLive

$$\frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} \sim t_{n-2}$$

Where t_{n-2} denotes the Student t-distribution with $n-2$ degrees of freedom. For our worked Melbourne household consumption example ($n = 6$, $SSR = 8$): $\hat{\sigma}^2 = 8/(6-2) = 2.0$, so $SE(\hat{\beta}_1) = \sqrt{2.0/4000} = 0.02236$. To test $H_0 : \beta_1 = 0$ against $H_1 : \beta_1 \neq 0$, the test statistic is $t = (0.70 - 0)/0.02236 = 31.30$, far exceeding the two-sided 5% critical value $t_{4,0.025} = 2.776$. A 95% confidence interval for the MPC is 0.70 ± 2.776(0.02236) = [0.638, 0.762].

1.7 Goodness of Fit: Variance Decomposition and R^2

To assess how closely the estimated sample regression line fits the observed sample data, total variation in y_i around its sample mean $\bar{y}$ is decomposed into explained and unexplained components. Writing $y_i - \bar{y} = (\hat{y}_i - \bar{y}) + \hat{u}_i$, squaring both sides, and summing across $i = 1, \ldots, n$ (noting the cross-product $\sum(\hat{y}_i - \bar{y})\hat{u}_i = 0$ by the OLS normal equations) produces:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}\hat{u}_i^2 \quad \Leftrightarrow \quad SST = SSE + SSR$$

Where:

- $SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$ is the Total Sum of Squares, measuring total sample variation in y.
- $SSE = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ is the Explained Sum of Squares, measuring variation in fitted values $\hat{y}_i$.
- $SSR = \sum_{i=1}^{n}\hat{u}_i^2$ is the Residual Sum of Squares, measuring unexplained variation.

The **Coefficient of Determination (R^2)** is defined as the fraction of total sample variation in y explained by x:

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

Where $R^2 \in [0, 1]$ whenever an intercept is included in the regression. In simple linear regression, R^2 equals the square of the sample correlation coefficient between y and x ($R^2 = r_{xy}^2$). In the Melbourne household consumption example, $SST = \sum(y_i - \bar{y})^2 = 1968$, and $SSR = 8$, so $R^2 = 1 - 8/1968 = 0.9959$ (99.59% of sample consumption variation is explained by income). Crucially, a high R^2 does not establish causality, nor does a low R^2 in microeconomic cross-sections (where individual behaviour is noisy) invalidate a statistically significant ceteris paribus effect.

DegreeLive

1.8 Data Scaling and Non-Linear Functional Forms

In applied econometrics (aligned with ULO3), analysts frequently rescale units of measurement or apply logarithmic and polynomial transformations to align regression models with economic theory, such as constant-elasticity demand curves or Mincer human-capital wage equations.

Effects of Linear Data Scaling

Suppose we rescale y_i by a positive constant c_1 (for example, converting Australian weekly earnings from dollars to thousands of dollars using $c_1 = 0.001$) and rescale x_i by a positive constant c_2 (for example, converting firm revenue from dollars to millions of dollars):

- If $y_i^* = c_1 y_i$ and x_i is unchanged, then $\hat{\beta}_0^* = c_1\hat{\beta}_0, \hat{\beta}_1^* = c_1\hat{\beta}_1$, and $SE(\hat{\beta}_1^*) = c_1 SE(\hat{\beta}_1)$.
- If $x_i^* = c_2 x_i$ and y_i is unchanged, then $\hat{\beta}_0^* = \hat{\beta}_0, \hat{\beta}_1^* = \hat{\beta}_1/c_2$, and $SE(\hat{\beta}_1^*) = SE(\hat{\beta}_1)/c_2$.
- In all linear rescalings of y or x, every t-statistic, p-value, and R^2 remains completely invariant because coefficients and their standard errors scale by the exact same factor.

Logarithmic and Quadratic Functional Forms

Because the linearity requirement SLR.1 restricts only the *parameters* to enter linearly, nonlinear transformations of y or x remain fully compatible with OLS estimation. Table 1.2 summarises the four primary functional specifications used in Australian empirical studies.

Model Name	Equation	Slope Interpretation (β_1)	Typical Economic Application
Level-Level	$y = \beta_0 + \beta_1 x + u$	$\Delta y = \beta_1 \Delta x$ (unit change)	Consumption vs. disposable income
Log-Level (Semi-log)	$\ln(y) = \beta_0 + \beta_1 x + u$	$\%\Delta y \approx (100 \cdot \beta_1)\Delta x$	Mincer wage equation: $\ln(\text{wage})$ on years of schooling
Level-Log	$y = \beta_0 + \beta_1 \ln(x) + u$	$\Delta y \approx (\beta_1/100)(\%\Delta x)$	Life satisfaction or health expenditure on log income
Log-Log (Double-log)	$\ln(y) = \beta_0 + \beta_1 \ln(x) + u$	$\%\Delta y \approx \beta_1(\%\Delta x)$ (elasticity)	Price elasticity of demand; CEO pay vs. ASX firm market cap

For the log-level specification $\ln(y_i) = \beta_0 + \beta_1 x_i + u_i$, the exact percentage change in predicted y for a discrete change Δx is $100 \cdot [\exp(\hat{\beta}_1 \Delta x) - 1]$, whereas $100\hat{\beta}_1 \Delta x$ is a first-order Taylor approximation accurate when $|\hat{\beta}_1 \Delta x|$ is small.

Worked Example (Australian Mincer Wage Equation & Quadratic Age Profile): Using a cross-section of $n = 1,200$ full-time Australian workers from ABS earnings data, suppose an analyst estimates a log-level regression of hourly wage (in AUD) on years of completed education (*educ*):

$$\widehat{\ln(wage)} = 2.15 + 0.082\, educ, \qquad R^2 = 0.241$$

Degree Live

Where $\ln(wage)$ is the natural logarithm of hourly wage in AUD and *educ* is completed years of formal education. Using the linear approximation, one additional year of schooling increases expected hourly wages by approximately 100 × 0.082 = 8.2%. Using the exact exponential formula, the predicted wage premium is 100 × [exp(0.082) − 1] = 8.55%. Notice that if wages were measured in cents instead of dollars ($wage^* = 100 \cdot wage$), then $\ln(wage^*) = \ln(100) + \ln(wage)$, shifting only the intercept from 2.15 to 2.15 + 4.605 = 6.755 while leaving the slope 0.082 unchanged—a major advantage of logarithmic dependent variables.

Similarly, diminishing marginal returns can be modelled with a quadratic term, such as Australian housing price (*price*, in AUD thousands) as a function of distance from the Melbourne CBD (*dist*, in km) or worker earnings as a function of labour-market experience (*exper*):

$$\widehat{wage} = 18.40 + 1.56\, exper - 0.026\, exper^2$$

Where *wage* is hourly wage in AUD and *exper* is years of workforce experience. Differentiating with respect to *exper* gives the marginal return to experience:

$$\frac{\partial \widehat{wage}}{\partial exper} = 1.56 - 0.052\, exper$$

Where $\partial \widehat{wage}/\partial exper$ is the marginal hourly wage increase per extra year of experience. For a worker with 10 years of experience, an additional year raises predicted wages by \$1.56 - 0.052(10) = 1.04($A1.04 per hour). Predicted wages reach their maximum turning point where the derivative equals zero: $exper^* = 1.56/0.052 = 30.0$ years, after which further experience exhibits a negative marginal slope.

Finally, when comparing a level model (y on x) against a log-level model ($\ln(y)$ on x), their raw R^2 values cannot be compared directly because their dependent variables have different total sums of squares ($\sum(y_i - \bar{y})^2$ versus $\sum(\ln y_i - \overline{\ln y})^2$). To compare fit properly, one exponentiates the fitted log values, applies the normality-corrected prediction $\widetilde{y_i} = \exp(\hat{\sigma}^2/2)\exp(\widehat{\ln y_i})$ (or Duan's smearing factor $n^{-1}\sum \exp(\hat{u}_i)$), and computes the squared correlation between y_i and $\widetilde{y_i}$.

DegreeLive