DegreeLive

Bachelor of Commerce (B2001) | Level 3 • Part B Major | Monash University

# Principles of econometrics

Course Code: ETC3400 • Module 2 Notes

# Module 2: Asymptotic OLS Theory, Generalised Least Squares & Hypothesis Testing

## 2.1 Modes of Stochastic Convergence and Asymptotic Limit Theorems

Exact finite-sample distribution theory relies heavily on strict exogeneity ($E(u \mid X) = 0$), spherical disturbances, and exact Gaussianity. In empirical microeconomics, finance, and macroeconomics—such as modelling Reserve Bank of Australia (RBA) policy reaction functions with lagged variables or ABS survey earnings with non-normal skewed wages—exact finite-sample distributions are generally unavailable. Asymptotic theory overcomes this limitation by characterising estimator behaviour as the sample size $n \to \infty$.

### Modes of Convergence for Random Sequences

Let $\{Z_n\}_{n=1}^{\infty}$ be a sequence of random vectors in $\mathrm{R}^k$ and $c \in \mathrm{R}^k$ a fixed vector.

1. **Convergence in Probability ($Z_n \xrightarrow{p} c$, or $\operatorname{plim} Z_n = c$):** For every $\varepsilon > 0$, $\lim_{n\to\infty} \Pr(\|Z_n - c\| > \varepsilon) = 0$. When an estimator $\hat{\theta}_n$ satisfies $\hat{\theta}_n \xrightarrow{p} \theta_0$, it is said to be **consistent**. A sufficient (though not necessary) condition is mean-square convergence: if $E(\hat{\theta}_n) \to \theta_0$ and $\operatorname{Var}(\hat{\theta}_n) \to 0$ as $n \to \infty$, then Chebyshev's inequality guarantees $\hat{\theta}_n \xrightarrow{p} \theta_0$.
2. **Convergence in Distribution ($Z_n \xrightarrow{d} Z$):** Let $F_n(z)$ be the cumulative distribution function (CDF) of $Z_n$ and $F(z)$ be the CDF of a limiting random vector $Z$. We say $Z_n$ converges in distribution to $Z$ if $\lim_{n\to\infty} F_n(z) = F(z)$ at every continuity point $z$ of $F$. Convergence in probability to a constant implies convergence in distribution to that constant, whereas the converse holds only when the limit is degenerate.

### Core Asymptotic Theorems and Operational Rules

| Theorem / Tool | Formal Statement | Primary Use in Econometrics |
|---|---|---|
| Weak Law of Large Numbers (Khinchin WLLN) | If $w_1, \ldots, w_n$ are i.i.d. with $E\|w_i\| < \infty$, then $\bar{w}_n = \frac{1}{n}\sum_{i=1}^{n} w_i \xrightarrow{p} E(w_i)$ | Establishing $\frac{1}{n}X'X \xrightarrow{p} Q_{xx}$ and $\frac{1}{n}X'u \xrightarrow{p} 0$ |
| Continuous Mapping Theorem (CMT) | If $Z_n \xrightarrow{p} c$ and $g(\cdot)$ is continuous at $c$, then $g(Z_n) \xrightarrow{p} g(c)$; similarly for $\xrightarrow{d}$ | Inverting $\hat{Q}_{xx}^{-1} \xrightarrow{p} Q_{xx}^{-1}$ and showing $g(\hat{\beta})$ is consistent |
| Lindeberg-Lévy Central Limit Theorem (CLT) | If $w_1, \ldots, w_n$ are i.i.d. with mean $0$ and finite covariance $\Sigma$, then $\underline{n}\bar{w}_n \xrightarrow{d} N(0, \Sigma)$ | Showing $\frac{1}{n}X'u \xrightarrow{d} N(0, \Omega)$ |
| Slutsky's Theorem | If $Z_n \xrightarrow{d} Z$ and $A_n \xrightarrow{p} A$, $b_n \xrightarrow{p} b$, then $A_n Z_n + b_n \xrightarrow{d} AZ + b$ | Combining $\hat{Q}_{xx}^{-1} \xrightarrow{p} Q_{xx}^{-1}$ with $\frac{1}{n}X'u \xrightarrow{d} N(0, \Omega)$ |

DegreeLive

| Theorem / Tool | Formal Statement | Primary Use in Econometrics |
|---|---|---|
| Delta Method | If $\underline{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, V)$ and $g$ is $C^1$ with Jacobian $G = \nabla g(\theta_0)'$, then $\underline{n}(g(\hat{\theta}_n) - g(\theta_0)) \xrightarrow{d} N(0, GVG')$ | Deriving standard errors for nonlinear functions of $\hat{\beta}$ (elasticities, ratios, marginal effects) |

### Stochastic Order Notation: $O_p(1)$ and $o_p(1)$

To streamline asymptotic proofs, a sequence $Z_n$ is $o_p(1)$ (small-oh-p one) if $Z_n \xrightarrow{p} 0$, and $O_p(1)$ (big-oh-p one, or bounded in probability) if for every $\varepsilon > 0$ there exists a finite $M_\varepsilon$ such that $\Pr(\|Z_n\| > M_\varepsilon) < \varepsilon$ for all sufficiently large $n$. Any sequence converging in distribution is automatically $O_p(1)$, and multiplying an $O_p(1)$ sequence by an $o_p(1)$ sequence yields $o_p(1)$. For example, $\hat{\beta} - \beta = O_p(n^{-1/2})$ indicates root-$n$ consistency.

## 2.2 Consistency and Asymptotic Normality of OLS

In large samples, we relax both strict exogeneity ($E(u \mid X) = 0$) and normality ($u \mid X \sim N(0, \sigma^2 I_n)$). Instead, consider a random sample (or stationary ergodic time series) $\{(y_i, x_i)\}_{i=1}^n$ generated by $y_i = x_i'\beta + u_i$ satisfying the following large-sample regularity conditions:

1. **Weak Exogeneity / Orthogonality:** $E(x_i u_i) = 0$. Notice this is strictly weaker than $E(u_i \mid x_i) = 0$ and much weaker than $E(u \mid X) = 0$, allowing lagged dependent variables in dynamic time-series regressions provided errors are serially uncorrelated with past information.
2. **Population Rank Identification:** $Q_{xx} = E(x_i x_i')$ is finite and strictly positive definite (hence invertible).
3. **Finite Fourth Moments:** $E\|x_i\|^4 < \infty$ and $E(u_i^4) < \infty$, ensuring that $\Omega = E(u_i^2 x_i x_i') = \mathrm{Var}(x_i u_i)$ is finite and positive definite.

### Proof of Consistency of OLS

Write the OLS sampling error by dividing both numerator and denominator by $n$:

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

Where:

- $\frac{1}{n}X'X = \frac{1}{n}\sum_{i=1}^n x_i x_i'$ is the sample second-moment matrix of regressors.
- $\frac{1}{n}X'u = \frac{1}{n}\sum_{i=1}^n x_i u_i$ is the sample cross-moment between regressors and errors.

By the Weak Law of Large Numbers, $\frac{1}{n}\sum_{i=1}^n x_i x_i' \xrightarrow{p} E(x_i x_i') = Q_{xx}$ and $\frac{1}{n}\sum_{i=1}^n x_i u_i \xrightarrow{p} E(x_i u_i) = 0$. Because matrix inversion $A \mapsto A^{-1}$ is continuous at every non-singular matrix $Q_{xx}$, the Continuous Mapping Theorem implies $\left(\frac{1}{n}X'X\right)^{-1} \xrightarrow{p} Q_{xx}^{-1}$. Applying Slutsky's multiplication rule yields:

$$\operatorname{plim}\hat{\beta} = \beta + Q_{xx}^{-1} \cdot 0 = \beta$$

Thus $\hat{\beta} \xrightarrow{p} \beta$: OLS is consistent even when $E(u \mid X) \neq 0$ or when disturbances are non-normal or heteroskedastic, provided $E(x_i u_i) = 0$.

### Proof of Asymptotic Normality of OLS

To derive the limiting distribution, scale the sampling error $\hat{\beta} - \beta$ by $\sqrt{n}$:

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

Where $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i u_i = \sqrt{n}\left(\frac{1}{n} \sum_{i=1}^{n} x_i u_i - 0\right)$. Since the vectors $z_i = x_i u_i$ are i.i.d. with mean $E(z_i) = 0$ and covariance matrix $\Omega = E(z_i z_i') = E(u_i^2 x_i x_i')$, the Multivariate Lindeberg-Lévy CLT guarantees:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i u_i \xrightarrow{d} N(0, \Omega)$$

Combining this convergence in distribution with $\left(\frac{1}{n} X'X\right)^{-1} \xrightarrow{p} Q_{xx}^{-1}$ via Slutsky's Theorem gives the general asymptotic distribution of OLS:

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N\left(0,\ Q_{xx}^{-1} \Omega Q_{xx}^{-1}\right)$$

Where:

- $Q_{xx}^{-1} \Omega Q_{xx}^{-1}$ is the asymptotic covariance matrix of $\sqrt{n}(\hat{\beta} - \beta)$, known as the **sandwich covariance matrix** because $\Omega$ (the "meat") is flanked by $Q_{xx}^{-1}$ (the "bread").
- Under conditional homoskedasticity $E(u_i^2 \mid x_i) = \sigma^2$, the meat simplifies to $\Omega = \sigma^2 E(x_i x_i') = \sigma^2 Q_{xx}$, so the sandwich collapses to $\sigma^2 Q_{xx}^{-1}$.

## 2.3 Heteroskedasticity-Consistent (HC) and HAC Covariance Estimation

When conditional error variance varies across observations—for example, across Australian households of different income levels or ASX firms of varying market capitalisation—homoskedastic OLS standard errors $\sqrt{s^2 [(X'X)^{-1}]_{jj}}$ are inconsistent, distorting $t$-statistics, $F$-tests, and confidence intervals even in arbitrarily large samples.

DegreeLive

## Eicker-Huber-White Heteroskedasticity-Consistent (HC) Estimator

To construct valid standard errors without specifying the functional form of heteroskedasticity $\sigma_i^2 = E(u_i^2 \mid x_i)$, White (1980) proposed replacing the unknown meat matrix $\Omega = E(u_i^2 x_i x_i')$ with its sample analogue formed from OLS residuals $\hat{u}_i = y_i - x_i'\hat{\beta}$:

$$\hat{\Omega}_{HC0} = \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2 x_i x_i' = \frac{1}{n} X' \hat{D} X, \qquad \hat{D} = \operatorname{diag}(\hat{u}_1^2, \hat{u}_2^2, \ldots, \hat{u}_n^2)$$

Where:

- $\hat{\Omega}_{HC0}$ converges in probability to $\Omega$ under finite fourth moments, despite individual $\hat{u}_i^2$ not consistently estimating individual $\sigma_i^2$.
- The resulting asymptotic variance estimator for $\hat{\beta}$ (scaled by $1/n$) is the White sandwich estimator:

$$\widehat{\operatorname{Var}}_{HC0}(\hat{\beta}) = (X'X)^{-1} \left(\sum_{i=1}^{n} \hat{u}_i^2 x_i x_i'\right) (X'X)^{-1}$$

Where $(X'X)^{-1}$ serves as the bread and $\sum_{i=1}^{n} \hat{u}_i^2 x_i x_i'$ serves as the meat.

Because $\hat{u}_i$ has smaller variance than $u_i$ in finite samples—specifically $\operatorname{Var}(\hat{u}_i \mid X) = (1 - h_i)\sigma_i^2$ under homoskedasticity, where $h_i = x_i'(X'X)^{-1}x_i$ is the $i$-th diagonal leverage of $P_X$—White's HC0 estimator is slightly downward-biased in small samples. Applied econometricians therefore use small-sample refinements:

- **HC1:** Multiplies $\widehat{\operatorname{Var}}_{HC0}(\hat{\beta})$ by the degrees-of-freedom factor $\frac{n}{n-k}$.
- **HC2:** Replaces $\hat{u}_i^2$ by $\hat{u}_i^2/(1 - h_i)$, correcting exact leverage bias under homoskedasticity.
- **HC3:** Replaces $\hat{u}_i^2$ by $\hat{u}_i^2/(1 - h_i)^2$, closely approximating a jackknife variance estimator and offering superior finite-sample coverage when high-leverage observations exist.

## Newey-West Heteroskedasticity and Autocorrelation Consistent (HAC) Estimator

In time-series regressions (such as quarterly Australian inflation or GDP equations), $x_t u_t$ may exhibit both heteroskedasticity and serial correlation across lags $j = 1, 2, \ldots$. Here the asymptotic variance of $\frac{1}{T}\sum_{t=1}^{T} x_t u_t$ becomes the long-run covariance matrix $\Omega_{LR} = \Gamma_0 + \sum_{j=1}^{\infty}(\Gamma_j + \Gamma_j')$, where $\Gamma_j = E(u_t u_{t-j} x_t x_{t-j}')$. Newey and West (1987) estimate $\Omega_{LR}$ consistently while guaranteeing positive semi-definiteness by tapering sample autocovariances with Bartlett kernel weights:

$$\hat{\Omega}_{HAC} = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t^2 x_t x_t' + \sum_{j=1}^{L}\left(1 - \frac{j}{L+1}\right)\frac{1}{T}\sum_{t=j+1}^{T} \hat{u}_t \hat{u}_{t-j}(x_t x_{t-j}' + x_{t-j} x_t')$$

DegreeLive

Where:

- $L$ is the truncation lag (bandwidth), chosen to grow slowly with sample size $T$ (commonly $L = \lfloor 4(T/100)^{2/9} \rfloor$ or $L = \lfloor 0.75T^{1/3} \rfloor$).
- $w_j = 1 - j/(L+1)$ is the Bartlett weight, decaying linearly from $1$ at lag $0$ to $0$ at lag $L+1$ to ensure $\hat{\Omega}_{HAC}$ remains positive semi-definite.

### Worked Example 1: Comparing Classical and White HC Standard Errors

Consider a simple cross-sectional regression without intercept $y_i = \beta x_i + u_i$ across $n = 4$ Australian retail outlets ($k = 1$), with $x = (1, 2, 3, 4)'$, $\hat{\beta} = 2.0$, and OLS residuals $\hat{u} = (-0.2, -0.4, -0.6, 0.7)'$ (verifying orthogonality: $\sum x_i \hat{u}_i = 1(-0.2) + 2(-0.4) + 3(-0.6) + 4(0.7) = 0$).

**Step 1: Compute classical homoskedastic standard error.**

First, $\sum_{i=1}^{4} x_i^2 = 1^2 + 2^2 + 3^2 + 4^2 = 30$, so $(X'X)^{-1} = 1/30$. The residual sum of squares is $\sum_{i=1}^{4} \hat{u}_i^2 = 0.04 + 0.16 + 0.36 + 0.49 = 1.05$. With $n - k = 3$ degrees of freedom, $s^2 = 1.05/3 = 0.35$. Thus:

$$\widehat{\mathrm{Var}}_{OLS}(\hat{\beta}) = \frac{s^2}{\sum x_i^2} = \frac{0.35}{30} \approx 0.01167 \implies \mathrm{se}_{OLS}(\hat{\beta}) = \sqrt{0.01167} \approx 0.1080$$

**Step 2: Compute White HC0 and HC1 robust standard errors.**

The squared weighted residuals $x_i^2 \hat{u}_i^2$ are $1(0.04) = 0.04$, $4(0.16) = 0.64$, $9(0.36) = 3.24$, and $16(0.49) = 7.84$, summing to $11.76$. Applying the sandwich formula gives:

$$\widehat{\mathrm{Var}}_{HC0}(\hat{\beta}) = \frac{\sum_{i=1}^{4} x_i^2 \hat{u}_i^2}{\left(\sum_{i=1}^{4} x_i^2\right)^2} = \frac{11.76}{30^2} = \frac{11.76}{900} \approx 0.01307 \implies \mathrm{se}_{HC0}(\hat{\beta}) \approx 0.1143$$

Applying the HC1 adjustment factor $n/(n-k) = 4/3$ yields $\widehat{\mathrm{Var}}_{HC1}(\hat{\beta}) = \frac{4}{3}(0.01307) = 0.01742$, so $\mathrm{se}_{HC1}(\hat{\beta}) \approx 0.1320$, which is $22.2\%$ larger than the classical standard error because residual dispersion rises sharply with $x_i$.

## 2.4 Generalised Least Squares (GLS) and Aitken's Theorem

While HC and HAC standard errors repair OLS inference under heteroskedasticity or autocorrelation, OLS itself is no longer BLUE when $\mathrm{Var}(u \mid X) \neq \sigma^2 I_n$. If the conditional covariance structure of the error vector is known up to a scalar multiple, efficiency can be restored via Generalised Least Squares (GLS).

Suppose Assumptions A1–A3 hold, but Assumption A4 is replaced by the non-spherical covariance specification:

DegreeLive

$$\text{Var}(u \mid X) = E(uu' \mid X) = \sigma^2\Omega$$

Where $\Omega$ is a known, symmetric, strictly positive definite $n \times n$ matrix ($\Omega > 0$). Because $\Omega$ is symmetric and positive definite, its inverse $\Omega^{-1}$ is also symmetric positive definite and admits a matrix square root $\Omega^{-1/2}$ (via spectral or Cholesky factorisation) satisfying $(\Omega^{-1/2})'\Omega^{-1/2} = \Omega^{-1}$ and $\Omega^{-1/2}\Omega(\Omega^{-1/2})' = I_n$.

Pre-multiplying the linear model $y = X\beta + u$ by $\Omega^{-1/2}$ transforms the variables into $y^* = \Omega^{-1/2}y$, $X^* = \Omega^{-1/2}X$, and $u^* = \Omega^{-1/2}u$, yielding the transformed regression:

$$y^* = X^*\beta + u^*$$

Where:

- $E(u^* \mid X^*) = \Omega^{-1/2}E(u \mid X) = 0$.
- $\text{Var}(u^* \mid X^*) = \Omega^{-1/2}\text{Var}(u \mid X)(\Omega^{-1/2})' = \sigma^2\Omega^{-1/2}\Omega(\Omega^{-1/2})' = \sigma^2 I_n$.
- $\text{rank}(X^*) = \text{rank}(\Omega^{-1/2}X) = \text{rank}(X) = k$ since $\Omega^{-1/2}$ is non-singular.

Because the transformed model $y^* = X^*\beta + u^*$ satisfies every Gauss-Markov assumption A1–A4, applying OLS to $(y^*, X^*)$ produces the **Generalised Least Squares (GLS) estimator**:

$$\hat{\beta}_{GLS} = (X^{*'}X^*)^{-1}X^{*'}y^* = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$$

Where:

- $\hat{\beta}_{GLS}$ minimises the generalised sum of squares $(y - Xb)'\Omega^{-1}(y - Xb)$.
- Conditional on $X$, $E(\hat{\beta}_{GLS} \mid X) = \beta$ and $\text{Var}(\hat{\beta}_{GLS} \mid X) = \sigma^2(X'\Omega^{-1}X)^{-1}$.

**Theorem (Aitken, 1935):** Under Assumptions A1–A3 and $\text{Var}(u \mid X) = \sigma^2\Omega$ with $\Omega > 0$ known, $\hat{\beta}_{GLS}$ is the Best Linear Unbiased Estimator (BLUE) of $\beta$. Specifically, comparing GLS against OLS ( $hat\beta_{OLS}$), whose true covariance matrix under $\sigma^2\Omega$ is $\text{Var}(\hat{\beta}_{OLS} \mid X) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$, we have:

$$\text{Var}(\hat{\beta}_{OLS} \mid X) - \text{Var}(\hat{\beta}_{GLS} \mid X) \geq 0$$

Where equality holds only in special cases such as $\Omega = I_n$ or when the columns of $\Omega X$ lie entirely inside the column space $\text{C}(X)$ (Kruskal's condition).

Degree Live

## 2.5 Weighted Least Squares (WLS) and Feasible GLS (FGLS)

An important special case of GLS arises under pure heteroskedasticity with uncorrelated observations, so $\Omega = \text{diag}(h_1, h_2, \dots, h_n)$ with $h_i > 0$. Here $\Omega^{-1/2} = \text{diag}(1/\sqrt{h_1}, \dots, 1/\sqrt{h_n})$, so the transformed model divides every observation's variables by $\sqrt{h_i}$:

$$\frac{y_i}{\sqrt{h_i}} = \frac{x_i'}{\sqrt{h_i}}\beta + \frac{u_i}{\sqrt{h_i}}$$

Applying OLS to this transformed equation defines the **Weighted Least Squares (WLS)** estimator, which minimises the weighted sum of squared residuals $\sum_{i=1}^{n} w_i(y_i - x_i'b)^2$ with weights $w_i = 1/h_i$:

$$\hat{\beta}_{WLS} = \left(\sum_{i=1}^{n} \frac{1}{h_i} x_i x_i'\right)^{-1} \left(\sum_{i=1}^{n} \frac{1}{h_i} x_i y_i\right)$$

Where observations with larger error variance $h_i$ receive smaller weight $w_i = 1/h_i$, preventing noisy observations from dominating parameter estimation.

### Feasible Generalised Least Squares (FGLS)

In empirical practice, $\Omega$ is rarely known a priori and must be estimated from the data using a parametric skedastic model $\sigma_i^2 = h(z_i, \gamma)$. A prominent specification is multiplicative heteroskedasticity, $\sigma_i^2 = \exp(z_i'\gamma)$, which guarantees strictly positive variance estimates. The **FGLS algorithm** proceeds in three steps:

1. Estimate the regression $y = X\beta + u$ by OLS and extract residuals $\hat{u}_i$.
2. Regress $\ln(\hat{u}_i^2)$ on $z_i$ (typically $x_i$ or powers of $x_i$) by OLS to obtain $\hat{\gamma}$, and construct predicted variances $\hat{h}_i = \exp(z_i'\hat{\gamma})$.
3. Compute the FGLS estimator $\hat{\beta}_{FGLS} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y$ via WLS with weights $\hat{w}_i = 1/\hat{h}_i$.

Because $\hat{\Omega}$ depends on the stochastic data $y$, $\hat{\beta}_{FGLS}$ is generally not unbiased in finite samples. However, provided $\hat{\gamma} \xrightarrow{p} \gamma$ and $E(u_i \mid x_i) = 0$, FGLS is consistent and asymptotically equivalent to infeasible GLS: $\sqrt{n}(\hat{\beta}_{FGLS} - \hat{\beta}_{GLS}) \xrightarrow{p} 0$, so $\sqrt{n}(\hat{\beta}_{FGLS} - \beta) \xrightarrow{d} N(0, [E(h_i^{-1}x_i x_i')]^{-1})$.

### Diagnostic Tests for Heteroskedasticity

- **Breusch-Pagan (BP) LM Test:** Under $H_0 : \text{Var}(u_i \mid x_i) = \sigma^2$, regress squared OLS residuals $\hat{u}_i^2$ on suspected variance drivers $z_i = (1, z_{i2}, \dots, z_{ip})'$ and compute $LM = nR^2_{\hat{u}^2} \xrightarrow{d} \chi^2_{p-1}$. Koenker's studentised version $nR^2_{\hat{u}^2}$ remains valid without normality.
- **White's General Test:** Regresses $\hat{u}_i^2$ on all regressors, their squares, and pairwise cross-products (or simply on $\hat{y}_i$ and $\hat{y}_i^2$ in the compact two-degree-of-freedom variant), evaluating $nR^2_{\hat{u}^2} \xrightarrow{d} \chi^2_q$.

DegreeLive

### Worked Example 2: Efficiency Comparison of OLS vs. WLS

Consider $y_i = \beta x_i + u_i$ across $n = 3$ firms with $x = (1, 2, 4)'$ and known heteroskedastic variances $\sigma_i^2 = \sigma^2 x_i^2$ (so $h_i = x_i^2$). Here $\sum_{i=1}^3 x_i^2 = 1 + 4 + 16 = 21$ and $\sum_{i=1}^3 x_i^4 = 1 + 16 + 256 = 273$.

Under OLS, the exact conditional variance of the slope estimator is:

$$\mathrm{Var}(\hat{\beta}_{OLS} \mid X) = \frac{\sum x_i^2 \sigma_i^2}{(\sum x_i^2)^2} = \sigma^2 \frac{\sum x_i^4}{(\sum x_i^2)^2} = \sigma^2 \frac{273}{21^2} = \sigma^2 \frac{273}{441} \approx 0.6190\sigma^2$$

Where $\sum x_i^2 = 21$ and $\sum x_i^4 = 273$. Dividing each equation by $\sqrt{h_i} = x_i$ gives $y_i/x_i = \beta + u_i/x_i$ with constant variance $\sigma^2$, so WLS yields $\hat{\beta}_{WLS} = \frac{1}{3}\sum_{i=1}^3 (y_i/x_i)$ with exact variance:

$$\mathrm{Var}(\hat{\beta}_{WLS} \mid X) = \frac{\sigma^2}{\sum_{i=1}^3 (x_i^2/x_i^2)} = \frac{\sigma^2}{3} \approx 0.3333\sigma^2$$

Where WLS achieves a $46.2\%$ variance reduction relative to OLS ($0.3333/0.6190 = 0.5385$).

## 2.6 Serial Correlation in Time Series: AR(1) Disturbances, Prais-Winsten & Testing

In macroeconomic and financial time-series data—such as quarterly Australian CPI inflation or ASX sector index returns—regression errors frequently exhibit serial dependence. Suppose $y_t = x_t'\beta + u_t$ ($t = 1, \ldots, T$) has disturbances following a stationary first-order autoregressive process, AR(1):

$$u_t = \rho u_{t-1} + \varepsilon_t, \qquad |\rho| < 1, \qquad \varepsilon_t \sim \text{i.i.d.}(0, \sigma_\varepsilon^2)$$

Where:

- $\rho$ is the first-order autocorrelation coefficient ($|\rho| < 1$ ensures covariance stationarity).
- $\varepsilon_t$ is the white-noise innovation process with unconditional error variance $\sigma_u^2 = \mathrm{Var}(u_t) = \sigma_\varepsilon^2/(1 - \rho^2)$ and autocovariance $\mathrm{Cov}(u_t, u_{t-s}) = \sigma_u^2 \rho^{|s|}$.

Collecting the autocovariances yields the Toeplitz correlation matrix $\Omega$ with $(t, s)$ entry $\rho^{|t-s|}/(1 - \rho^2)$. Inverting $\Omega$ directly produces a tridiagonal precision matrix whose Cholesky factor $\Omega^{-1/2}$ yields the exact **Prais-Winsten GLS transformation** across all $T$ observations:

$$y_1^* = \sqrt{1 - \rho^2}\, y_1, \qquad x_1^* = \sqrt{1 - \rho^2}\, x_1$$

DegreeLive

$$y_t^* = y_t - \rho y_{t-1}, \qquad x_t^* = x_t - \rho x_{t-1} \qquad (t = 2, \dots, T)$$

Where:

- For $t \geq 2$, quasi-differencing produces $u_t^* = u_t - \rho u_{t-1} = \varepsilon_t$ with variance $\sigma_\varepsilon^2$.
- For $t = 1$, scaling by $\sqrt{1-\rho^2}$ gives $\text{Var}(\sqrt{1-\rho^2}\, u_1) = (1-\rho^2)\sigma_u^2 = \sigma_\varepsilon^2$, preserving the initial observation.
- By contrast, the **Cochrane-Orcutt** procedure omits $t = 1$ and applies OLS only to $t = 2, \dots, T$; although asymptotically equivalent to Prais-Winsten, dropping $t = 1$ loses efficiency in short macroeconomic samples when regressors vary smoothly over time.

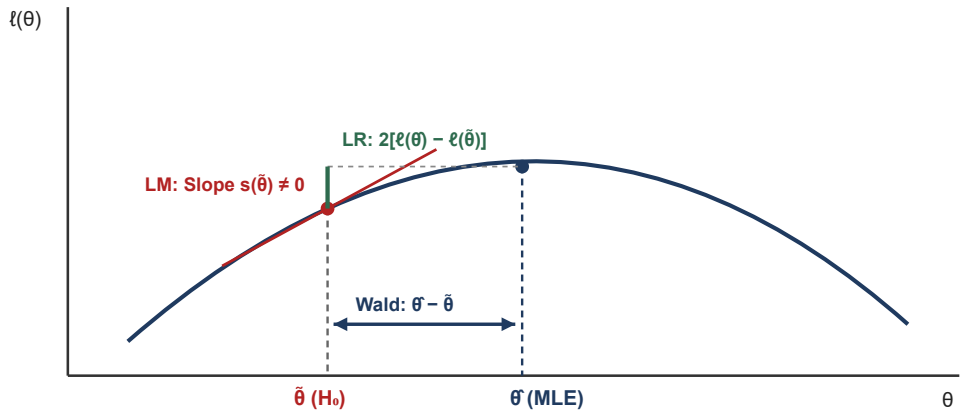
### Testing for Serial Correlation

When $\rho$ is unknown, FGLS estimates $\hat{\rho}$ from OLS residuals $\hat{u}_t$ via $\hat{\rho} = \sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1} / \sum_{t=1}^{T-1} \hat{u}_t^2$ (and iterates until convergence). Before applying FGLS or HAC standard errors, analysts test $H_0 : \rho = 0$:

- **Durbin-Watson (DW) Statistic:** $DW = \sum_{t=2}^{T} (\hat{u}_t - \hat{u}_{t-1})^2 / \sum_{t=1}^{T} \hat{u}_t^2 \approx 2(1 - \hat{\rho})$. Values near $2$ indicate no serial correlation; values below $d_L$ signal positive autocorrelation. Critically, DW is invalid whenever $X$ includes lagged dependent variables $y_{t-1}$ or lacks strict exogeneity.
- **Breusch-Godfrey (BG) LM Test:** Valid even with lagged dependent variables and higher-order AR($p$) errors ($u_t = \rho_1 u_{t-1} + \cdots + \rho_p u_{t-p} + \varepsilon_t$). Regressing $\hat{u}_t$ on $x_t$ and $\hat{u}_{t-1}, \dots, \hat{u}_{t-p}$ (setting pre-sample residuals to $0$) yields $LM = (T-p)R_{\hat{u}}^2 \overset{d}{\sim} \chi_p^2$ (or $TR_{\hat{u}}^2 \overset{d}{\sim} \chi_p^2$) under $H_0 : \rho_1 = \cdots = \rho_p = 0$.

## 2.7 The Classical Hypothesis Testing Trinity: Wald, LM, and LR Tests

Beyond linear regression, econometric models frequently test $q$ potentially nonlinear restrictions $H_0 : g(\theta) = 0$ on a $k \times 1$ parameter vector $\theta$ estimated by Maximum Likelihood (or GMM), where $g : \mathrm{R}^k \to \mathrm{R}^q$ has full-rank Jacobian $G(\theta) = \partial g(\theta)/\partial \theta'$ $(q \times k)$. Three asymptotically equivalent likelihood-based test principles form the classical testing trinity: the **Wald ($W$)**, **Likelihood Ratio ($LR$)**, and **Lagrange Multiplier ($LM$, or Score)** tests.



1. **Wald Test ($W$ — Requires Only Unrestricted Estimate $\hat{\theta}$):** Measures how far the unrestricted estimate $g(\hat{\theta})$ lies from $0$, weighted by the inverse of its estimated asymptotic covariance matrix. Using the Delta Method, $\sqrt{n}(g(\hat{\theta}) - g(\theta_0)) \xrightarrow{d} N(0, G(\theta_0)V(\theta_0)G(\theta_0)')$, where $\widehat{\text{Var}}(\hat{\theta}) = \frac{1}{n}\hat{V}$:

$$W = g(\hat{\theta})' \left[G(\hat{\theta})\, \widehat{\text{Var}}(\hat{\theta})\, G(\hat{\theta})'\right]^{-1} g(\hat{\theta}) \xrightarrow{d} \chi^2_q$$

   Where $W$ is particularly convenient when the restricted model is nonlinear or difficult to estimate.

2. **Likelihood Ratio Test ($LR$ — Requires Both $\hat{\theta}$ and $\tilde{\theta}$):** Compares the maximised unrestricted log-likelihood $\ell(\hat{\theta})$ against the maximised restricted log-likelihood $\ell(\tilde{\theta})$ subject to $g(\tilde{\theta}) = 0$:

$$LR = 2\left[\ell(\hat{\theta}) - \ell(\tilde{\theta})\right] \xrightarrow{d} \chi^2_q$$

   Where $LR$ is invariant to nonlinear reparameterisations of $g(\theta) = 0$, unlike the Wald statistic.

3. **Lagrange Multiplier / Score Test ($LM$ — Requires Only Restricted Estimate $\tilde{\theta}$):** Evaluates the gradient (score vector) $s(\theta) = \partial\ell(\theta)/\partial\theta$ at the restricted estimate $\tilde{\theta}$. Because $\ell(\theta)$ peaks at $\hat{\theta}$ where $s(\hat{\theta}) = 0$, if $H_0$ is true then $\tilde{\theta}$ lies close to $\hat{\theta}$ and $s(\tilde{\theta})$ should be close to $0$:

$$LM = s(\tilde{\theta})'\, \mathrm{I}(\tilde{\theta})^{-1}\, s(\tilde{\theta}) \xrightarrow{d} \chi^2_q$$

   Where $\mathrm{I}(\tilde{\theta})$ is the Fisher information matrix evaluated at $\tilde{\theta}$. In OLS, $LM$ often simplifies to $nR^2$ from an auxiliary residual regression (as in the Breusch-Pagan and Breusch-Godfrey tests).

## Asymptotic Equivalence and Finite-Sample Inequality

Under $H_0$ and standard regularity conditions (proved via second-order Taylor expansions of $\ell(\theta)$ around $\hat{\theta}$), $W$, $LR$, and $LM$ converge in probability to one another ($W - LR \xrightarrow{p} 0$ and $LR - LM \xrightarrow{p} 0$) and share the same limiting $\chi^2_q$ null distribution as well as identical local asymptotic power. In the classical linear normal regression model with $q$ linear restrictions $R\beta = r$, Savin (1976) and Berndt and Savin (1977) proved the exact deterministic inequality:

$$W = n\left(\frac{SSR_R - SSR_U}{SSR_U}\right) \geq LR = n\ln\left(\frac{SSR_R}{SSR_U}\right) \geq LM = n\left(\frac{SSR_R - SSR_U}{SSR_R}\right)$$

DegreeLive

Where $SSR_U$ and $SSR_R$ are the unrestricted and restricted residual sums of squares, respectively. Note that for linear models, $W = qF \cdot \frac{n}{n-k}$, showing why finite-sample Wald tests using $\chi^2_q$ critical values slightly over-reject in small samples relative to exact $F_{q,n-k}$ tests.

## Worked Example 3: Computing $W$, $LR$, and $LM$ for Linear Restrictions

Suppose an econometrician estimates a linear model with $n = 100$ observations and $k = 4$ parameters, testing $q = 2$ linear restrictions $H_0 : R\beta = r$. Unrestricted and restricted OLS yield $SSR_U = 80.0$ and $SSR_R = 92.0$.

Computing the three test statistics under Gaussian maximum likelihood gives:

- **Wald Statistic:** $W = 100 \times \frac{92.0-80.0}{80.0} = 100 \times 0.15 = 15.00$.
- **Likelihood Ratio Statistic:** $LR = 100 \ln(\frac{92.0}{80.0}) = 100 \ln(1.15) \approx 13.98$.
- **Lagrange Multiplier Statistic:** $LM = 100 \times \frac{92.0-80.0}{92.0} = 100 \times 0.130435 \approx 13.04$.

The exact ordering $W = 15.00 \geq LR = 13.98 \geq LM = 13.04$ holds identically. Comparing all three against the $5\%$ $\chi^2_2$ critical value of $5.99$ leads to decisive rejection of $H_0$ under every member of the testing trinity.

DegreeLive