DegreeLive

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

Introductory econometrics

Course Code: ETC2410 • Module 4 Notes

Module 4: Time-Series Econometrics, Panel Data Foundations & Limited Dependent Variables

4.1 Stationarity, Weak Dependence, and Random Walks

In cross-sectional analysis, random sampling ensures observations are mutually independent. By contrast, macroeconomic and financial time series—such as quarterly Australian real GDP, the ASX 200 index, or the RBA cash rate—exhibit persistence across time. Evaluating whether OLS estimators remain consistent and whether standard inferential distributions remain valid requires formalising the concepts of **stationarity** and **integration**.

A stochastic process $\{y_t\}_{t=1}^{\infty}$ is **covariance (weakly) stationary** if its first two moments are invariant to shifts in time:

1. $E(y_t) = \mu$ (constant finite mean for all t).
2. $\text{Var}(y_t) = \gamma_0 < \infty$ (constant finite variance for all t).
3. $\text{Cov}(y_t, y_{t-h}) = \gamma_h$ depends solely on the lag separation $h \geq 0$, not on the calendar date t.

If, additionally, $\gamma_h \to 0$ sufficiently rapidly as $h \to \infty$, the series is **weakly dependent**, enabling Laws of Large Numbers and Central Limit Theorems to hold for time-series averages. An important stationary, weakly dependent benchmark is the stable first-order autoregressive process **AR(1)**:

$$y_t = c + \rho y_{t-1} + e_t, \qquad |\rho| < 1$$

Where:

- $e_t \sim \text{WN}(0, \sigma_e^2)$ is a white-noise innovation (zero mean, constant variance σ_e^2, zero autocorrelation).
- $E(y_t) = \mu = c/(1-\rho)$ is the unconditional long-run mean.
- $\text{Var}(y_t) = \sigma_e^2/(1-\rho^2)$ is the finite unconditional variance, and $\text{Corr}(y_t, y_{t-h}) = \rho^h \to 0$ geometrically.

When $\rho = 1$, the process becomes a **unit root** or **Random Walk** (with drift if $c \neq 0$):

$$y_t = c + y_{t-1} + e_t \quad \Longrightarrow \quad y_t = y_0 + ct + \sum_{s=1}^{t} e_s$$

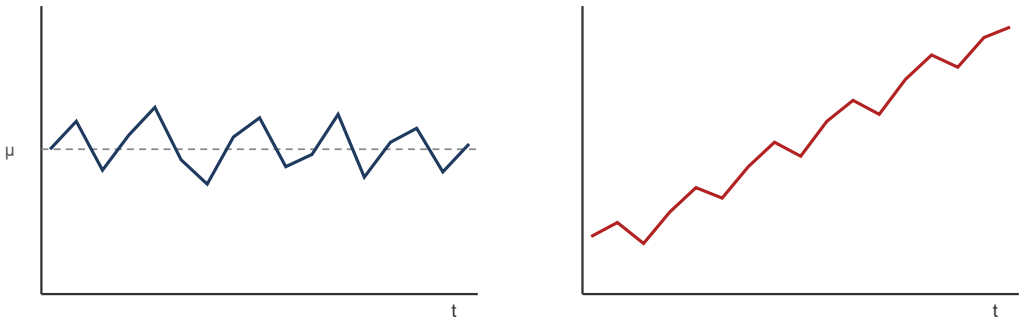
Where:

- y_0 is the initial value at $t = 0$.
- $\sum_{s=1}^{t} e_s$ is the cumulative sum of all past shocks; every innovation persists permanently rather than decaying.

DegreeLive

- Even when $c = 0$, $\text{Var}(y_t \mid y_0) = t\sigma_e^2$, which grows linearly without bound as $t \to \infty$, violating covariance stationarity.

A stationary, weakly dependent series is called **integrated of order zero**, denoted $I(0)$. A non-stationary series that becomes stationary after taking d first differences ($\Delta y_t = y_t - y_{t-1}$) is **integrated of order** d, denoted $I(d)$. Because the first difference of a random walk with drift is $\Delta y_t = c + e_t$, which is stationary white noise around c, a random walk is $I(1)$.



(a) Stationary I(0): Mean-Reverting AR(1)

(b) Non-Stationary I(1): Random Walk

4.2 Unit Root Testing and Spurious Regression

Regressing one non-stationary $I(1)$ variable on an independent non-stationary $I(1)$ variable in levels creates the **spurious regression problem** (Granger & Newbold, 1974). Even when two random walks are generated completely independently, OLS frequently produces a deceptively high R^2, large $|t|$-statistics that falsely reject the null of zero relationship, and a very low Durbin-Watson statistic ($DW \approx 0$, often satisfying the rule-of-thumb warning sign $R^2 > DW$). Because the error term itself is $I(1)$, standard t and F distributions diverge rather than converging to standard normal or t limits.

To test formally whether a series contains a unit root ($I(1)$) versus being stationary ($I(0)$), econometricians apply the **Augmented Dickey-Fuller (ADF) test**. Subtracting y_{t-1} from an $AR(p)$ process with intercept and optional linear trend gives:

$$\Delta y_t = \alpha + \delta t + \theta y_{t-1} + \sum_{j=1}^{p-1} \gamma_j \Delta y_{t-j} + e_t$$

Where:

- $\theta = \rho - 1$ is the unit-root parameter of interest.
- α is a constant drift term and δt is an optional deterministic time trend (included when the series exhibits a trending trajectory, such as log Australian real GDP).
- $\sum_{j=1}^{p-1} \gamma_j \Delta y_{t-j}$ adds lagged first differences to ensure e_t is serially uncorrelated white noise.

The null hypothesis is $H_0 : \theta = 0$ (unit root / non-stationary $I(1)$) against the one-sided stationary alternative $H_1 : \theta < 0$. Because y_{t-1} is non-stationary under H_0, the t-ratio $\tau = \hat{\theta}/SE(\hat{\theta})$ does *not* follow a standard Student t-distribution; instead, it is compared against non-standard **Dickey-Fuller**

critical values (for example, at 5% with $T = 100$: -2.89 with intercept only, and -3.45 with intercept and trend). If we fail to reject H_0 in levels, we test Δy_t to confirm it is $I(0)$ before proceeding with differenced regressions—unless the levels are cointegrated.

Worked ADF Test Calculation (Australian Real GDP): Suppose an economist tests quarterly log Australian real GDP ($lgdp_t$) over $T = 120$ quarters using an ADF specification with an intercept, linear trend, and one lagged difference ($p = 2$). The OLS regression outputs $\Delta lgdp_t = 0.24 + 0.0004t - 0.032\, lgdp_{t-1} + 0.28\, \Delta lgdp_{t-1}$, with $SE(\hat{\theta}) = 0.016$. Computing the Dickey-Fuller test statistic gives $\tau = -0.032/0.016 = -2.00$. Comparing against the 5% trend-included critical value (-3.45), because $-2.00 > -3.45$ (less negative than the critical cutoff), we fail to reject H_0 and conclude $lgdp_t$ contains a unit root. Re-running the ADF test on $\Delta lgdp_t$ (with intercept only) yields $\hat{\theta} = -0.74$ with $SE = 0.11$, so $\tau = -6.73 < -2.89$: quarterly GDP growth is stationary, confirming $lgdp_t \sim I(1)$.

Deterministic Trends, Detrending, and Quarterly Seasonality

Many Australian macroeconomic and business series grow steadily over time due to population growth, capital accumulation, or technological progress, or fluctuate predictably across calendar quarters (such as Q4 pre-Christmas retail sales). A **trend-stationary** process combines a deterministic trend with stationary fluctuations, either linear ($y_t = \alpha_0 + \alpha_1 t + e_t$) or exponential ($\ln y_t = \beta_0 + \beta_1 t + e_t$, where β_1 is the constant per-period growth rate). Including a linear time trend t directly in a multiple regression $y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \gamma t + u_t$ is equivalent, by the Frisch-Waugh-Lovell Theorem, to first detrending y_t and every regressor separately and then regressing detrended y_t on detrended x_{t1} and x_{t2}. This prevents spurious correlations driven solely by shared secular trends.

Similarly, to capture deterministic seasonality in quarterly Australian data without losing observations, include three seasonal dummy variables (for example, $Q2_t, Q3_t, Q4_t$, leaving Quarter 1 as the reference season) alongside an intercept:

$$y_t = \beta_0 + \delta_2 Q2_t + \delta_3 Q3_t + \delta_4 Q4_t + \beta_1 x_{t1} + \gamma t + u_t$$

Where $\delta_2, \delta_3, \delta_4$ measure seasonal shifts relative to Quarter 1, and joint significance of seasonality is assessed via an F-test of $H_0 : \delta_2 = \delta_3 = \delta_4 = 0$ ($q = 3$). Crucially, first-differencing a $I(1)$ random-walk-with-drift process removes stochastic trends, whereas linear detrending removes only deterministic trends; applying deterministic detrending to an $I(1)$ unit-root series leaves a non-stationary stochastic trend intact.

4.3 Dynamic Time-Series Models: FDL, AR(p), and ADL(p, q)

Economic adjustments unfold over multiple quarters rather than instantaneously. For instance, an RBA cash rate change takes several quarters to pass fully through to Australian mortgage rates, business investment, and inflation. Three workhorse dynamic models capture these lags:

- **Finite Distributed Lag Model ($FDL(q)$):** Models y_t as a function of current and q lagged values of an explanatory variable x_t:

degreelive

$$y_t = \alpha_0 + \delta_0 x_t + \delta_1 x_{t-1} + \cdots + \delta_q x_{t-q} + u_t$$

Where δ_0 is the **impact (short-run) propensity** (immediate change in y_t from a temporary one-unit shock to x_t) and $\text{LRP} = \sum_{j=0}^{q} \delta_j$ is the **long-run propensity (cumulative multiplier)** after a permanent one-unit increase in x.

- **Autoregressive Model ($AR(p)$):** Forecasts a stationary series solely from its own past history: $y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + e_t$, with lag order p selected via AIC or BIC.
- **Autoregressive Distributed Lag Model ($ADL(p,q)$):** Combines both lagged dependent variables and distributed lags of x_t:

$$y_t = \alpha_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=0}^{q} \delta_j x_{t-j} + e_t$$

Where e_t is serially uncorrelated provided sufficient lags (p,q) are included. Under dynamic completeness ($E(e_t \mid y_{t-1}, x_t, x_{t-1}, \dots) = 0$) and stationarity, OLS is consistent and asymptotically normal. The long-run equilibrium multiplier in a stable $ADL(1,1)$ model $y_t = \alpha_0 + \phi_1 y_{t-1} + \delta_0 x_t + \delta_1 x_{t-1} + e_t$ (with $|\phi_1| < 1$) is $\theta = (\delta_0 + \delta_1)/(1 - \phi_1)$.

Dynamic Completeness and the Expectations-Augmented Phillips Curve

A critical distinction in time-series regression separates **static models** ($y_t = \beta_0 + \beta_1 x_t + u_t$) from **dynamically complete models** where enough lags of y_t and x_t appear so that $E(u_t \mid \mathbf{x}_t, y_{t-1}, \mathbf{x}_{t-1}, \dots) = 0$. Under dynamic completeness, the error term u_t is automatically serially uncorrelated, making OLS consistent and conventional or White standard errors valid without HAC adjustments. For instance, in Australian macroeconomics, a static Phillips curve $inf_t = \beta_0 + \beta_1 unemp_t + u_t$ omits inflation expectations ($inf_t^e \approx inf_{t-1}$), inducing strong serial correlation. Reformulating as an expectations-augmented dynamic model:

$$\Delta inf_t = \alpha_0 + \phi_1 \Delta inf_{t-1} + \beta_1 (unemp_t - u^*) + e_t$$

Where $\Delta inf_t = inf_t - inf_{t-1}$ is the quarterly change in annual CPI inflation, u^* is the Non-Accelerating Inflation Rate of Unemployment (NAIRU), and $\beta_1 < 0$ captures the short-run inflation-unemployment trade-off. If the RBA estimates $\widehat{\Delta inf_t} = 1.90 - 0.42\, unemp_t$ (omitting lags for illustration), the implied NAIRU where $\Delta inf_t = 0$ is $u^* = 1.90/0.42 = 4.52\%$.

Out-of-Sample Forecast Evaluation

When dynamic $AR(p)$ or $ADL(p,q)$ models are used by institutions such as the RBA or Australian Treasury to forecast inflation or unemployment, model selection relies on out-of-sample forecast accuracy rather than in-sample R^2 (which rewards overfitting). Splitting the sample at T_0 into an

Degree.live

estimation window ($t = 1, \dots, T_0$) and an evaluation window ($t = T_0 + 1, \dots, T_0 + m$) with forecast errors $e_{t+1|t} = y_{t+1} - \hat{y}_{t+1|t}$, standard accuracy metrics are:

$$\text{RMSFE} = \sqrt{\frac{1}{m}\sum_{h=1}^{m}(y_{T_0+h} - \hat{y}_{T_0+h|T_0+h-1})^2}, \qquad \text{MAE} = \frac{1}{m}\sum_{h=1}^{m}|y_{T_0+h} - \hat{y}_{T_0+h|T_0+h-1}|$$

Where RMSFE (Root Mean Squared Forecast Error) penalises large forecast misses more severely than MAE (Mean Absolute Error) and aligns with the squared-error loss function implied by conditional mean forecasting.

4.4 Cointegration and Error Correction Models (ECM)

While regressing two unrelated $I(1)$ series yields a spurious regression, two or more $I(1)$ variables may share a common stochastic trend driven by a stable long-run economic equilibrium. If y_t and x_t are both $I(1)$, they are **cointegrated** if there exists a scalar β such that the linear combination $z_t = y_t - \beta x_t$ is stationary, $I(0)$.

Classic Australian macroeconomic examples include household consumption and disposable income (which drift upward together over decades while maintaining a stable long-run savings ratio), short-term and long-term Commonwealth Government bond yields, and domestic prices relative to foreign prices adjusted for the AUD exchange rate. In the **Engle-Granger two-step procedure**:

1. Estimate the long-run equilibrium regression $y_t = \beta_0 + \beta_1 x_t + z_t$ by OLS and save residuals $\hat{z}_t = y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t$.
2. Perform an ADF unit-root test on $\hat{z}_t$ (without trend, using Engle-Granger critical values, which are stricter than standard DF tables—approximately -3.37 at 5% for two variables—because $\hat{\beta}_1$ was estimated). Rejecting a unit root in $\hat{z}_t$ confirms cointegration.

By the Granger Representation Theorem, cointegrated variables admit an **Error Correction Model (ECM)** linking short-run changes to deviations from the long-run equilibrium:

$$\Delta y_t = \alpha_0 + \gamma_0 \Delta x_t + \lambda(y_{t-1} - \hat{\beta}_0 - \hat{\beta}_1 x_{t-1}) + e_t, \qquad -1 < \lambda < 0$$

Where:

- γ_0 measures the short-run contemporaneous impact of Δx_t on Δy_t.
- $\hat{z}_{t-1} = y_{t-1} - \hat{\beta}_0 - \hat{\beta}_1 x_{t-1}$ is the lagged equilibrium error (disequilibrium gap).
- $\lambda < 0$ is the **speed-of-adjustment coefficient**: when y_{t-1} overshoots its long-run equilibrium ($\hat{z}_{t-1} > 0$), $\lambda \hat{z}_{t-1} < 0$ pulls Δy_t downward back toward equilibrium.

Worked Numerical ECM Example (Australian Consumption-Income Equilibrium): Using quarterly ABS national accounts from 1995 to 2025 ($T = 120$), log real consumption (c_t) and log real disposable income (inc_t) are both $I(1)$. Step 1 estimates the cointegrating relation $\hat{c}_t = 0.12 +$

$0.94\,inc_t$, and the Engle-Granger ADF statistic on $\hat{z}_t$ is $-4.12 < -3.37$, confirming cointegration. Step 2 estimates the ECM:

$$\Delta \widehat{c}_t = 0.002 + 0.42\,\Delta inc_t - 0.25\,\hat{z}_{t-1}$$

Here, a 1% short-run rise in disposable income raises quarterly consumption immediately by 0.42% (short-run elasticity), compared with the long-run elasticity of 0.94. If consumption in quarter $t-1$ was 2% above its long-run equilibrium ($\hat{z}_{t-1} = +0.02$), the error-correction mechanism adjusts consumption downward by $-0.25(0.02) = -0.005$ (-0.5%) in the following quarter, closing 25% of any disequilibrium gap per quarter.

4.5 Panel Data Foundations: Pooled OLS, Fixed Effects & Random Effects

Panel datasets track the same cross-sectional units $i = 1, \dots, N$ over time periods $t = 1, \dots, T$ (such as Australian workers in HILDA, ASX-listed firms, or the eight Australian states and territories). A linear unobserved-effects panel model is written:

$$y_{it} = \beta_0 + \mathbf{x}_{it}'\boldsymbol{\beta} + a_i + u_{it}$$

Where:

- $\mathbf{x}_{it}$ is a $k \times 1$ vector of time-varying explanatory variables for unit i in period t.
- a_i is the **unobserved time-invariant heterogeneity** (individual fixed effect), absorbing constant attributes such as innate worker ability, corporate culture, or geographic climate.
- u_{it} is the idiosyncratic time-varying error term.

Because a_i is unobserved, estimating the model by **Pooled OLS** leaves the composite error $v_{it} = a_i + u_{it}$ in the disturbance. If $\text{Cov}(a_i, \mathbf{x}_{it}) \neq 0$ (for example, higher ability a_i increases both education/training and wages), Pooled OLS suffers from severe omitted variable bias. Two estimators eliminate a_i without requiring it to be measured:

1. **First-Differenced (FD) Estimator:** Subtracting period $t-1$ from period t cancels a_i directly:

   $$\Delta y_{it} = \Delta \mathbf{x}_{it}'\boldsymbol{\beta} + \Delta u_{it}, \qquad t = 2, \dots, T$$

   Where $\Delta y_{it} = y_{it} - y_{i,t-1}$. Under strict exogeneity $E(u_{it} \mid \mathbf{x}_{i1}, \dots, \mathbf{x}_{iT}, a_i) = 0$, OLS on first differences is unbiased and consistent even when a_i is arbitrarily correlated with $\mathbf{x}_{it}$.
2. **Fixed Effects (FE / Within) Estimator:** Subtracting each unit's time mean $\bar{y}_i = T^{-1}\sum_{t=1}^{T} y_{it}$ from y_{it} also eliminates a_i (since $\bar{a}_i = a_i$):

DegreeLive

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\boldsymbol{\beta} + (u_{it} - \bar{u}_i)$$

Applying OLS to these time-demeaned variables yields $\hat{\boldsymbol{\beta}}_{FE}$. Equivalently, FE equals OLS with N unit-specific dummy intercepts (Least Squares Dummy Variables). Note that any strictly time-invariant variable (such as sex or country of birth) is swept away by both FD and FE transformations, so its coefficient cannot be identified separately from a_i.

Alternatively, if the analyst is confident that unobserved heterogeneity is completely uncorrelated with all explanatory variables, $\text{Cov}(a_i, \mathbf{x}_{it}) = \mathbf{0}$, then Pooled OLS is consistent but inefficient because $v_{it} = a_i + u_{it}$ is serially correlated across time within the same unit ($\text{Corr}(v_{it}, v_{is}) = \sigma_a^2/(\sigma_a^2 + \sigma_u^2) > 0$). Under $\text{Cov}(a_i, \mathbf{x}_{it}) = \mathbf{0}$, the **Random Effects (RE)** estimator applies FGLS quasi-time-demeaning:

$$y_{it} - \theta\bar{y}_i = (1-\theta)\beta_0 + (\mathbf{x}_{it} - \theta\bar{\mathbf{x}}_i)'\boldsymbol{\beta} + (v_{it} - \theta\bar{v}_i), \qquad \theta = 1 - \sqrt{\frac{\sigma_u^2}{\sigma_u^2 + T\sigma_a^2}}$$

Where $\theta \in [0, 1]$ bridges Pooled OLS ($\theta = 0$ when $\sigma_a^2 = 0$) and Fixed Effects ($\theta \to 1$ as T or σ_a^2/σ_u^2 grows large). Crucially, RE allows estimation of time-invariant covariates, but becomes inconsistent if $\text{Cov}(a_i, \mathbf{x}_{it}) \neq \mathbf{0}$.

Estimator	Key Identifying Assumption	Allows $\text{Cov}(a_i, \mathbf{x}_{it}) \neq 0$?	Estimates Time-Invariant z_i?
Pooled OLS	$\text{Cov}(a_i + u_{it}, \mathbf{x}_{it}) = 0$	No (Biased & Inconsistent)	Yes
First Differences (FD)	$E(u_{it} \mid \mathbf{X}_i, a_i) = 0$	Yes (Consistent)	No (Differenced out)
Fixed Effects (FE)	$E(u_{it} \mid \mathbf{X}_i, a_i) = 0$	Yes (Consistent; more efficient than FD if u_{it} i.i.d.)	No (Demeaned out)
Random Effects (RE)	$E(u_{it} \mid \mathbf{X}_i, a_i) = 0$ & $\text{Cov}(a_i, \mathbf{x}_{it}) = 0$	No (Inconsistent)	Yes

Choosing Between FE and RE: The Hausman Test

To test $H_0 : \text{Cov}(a_i, \mathbf{x}_{it}) = \mathbf{0}$ (under which both FE and RE are consistent, and RE is efficient) against $H_1 : \text{Cov}(a_i, \mathbf{x}_{it}) \neq \mathbf{0}$ (under which FE remains consistent while RE is inconsistent), the **Hausman specification test** compares the difference vector $\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE}$:

$$H = (\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE})' \left[\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_{FE}) - \widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_{RE})\right]^{-1} (\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE}) \overset{a}{\sim} \chi_k^2 \quad \text{under } H_0$$

Degreelive

Where k is the number of time-varying regressors compared. Rejecting H_0 indicates systematic divergence between FE and RE coefficients, dictating Fixed Effects. In panel inference, standard errors should routinely be clustered at the cross-sectional unit level (i) to accommodate arbitrary serial correlation and heteroskedasticity within each unit across time.

Worked Numerical Panel Example: Pooled OLS vs. FD vs. Fixed Effects

Consider a balanced panel of $N = 3$ Australian manufacturing firms observed over $T = 2$ financial years, recording training hours per worker (x_{it}) and labour productivity index (y_{it}). Suppose unobserved managerial quality a_i varies across firms ($a_1 < a_2 < a_3$) and correlates positively with training.

Firm (i)	Year (t)	x_{it}	y_{it}	Δx_i ($t = 2 - 1$)	Δy_i ($t = 2 - 1$)	$x_{it} - \bar{x}_i$	$y_{it} - \bar{y}_i$
1	1	2	10	—	—	−1.0	−1.5
1	2	4	13	2	3	+1.0	+1.5
2	1	5	18	—	—	−1.5	−2.0
2	2	8	22	3	4	+1.5	+2.0
3	1	10	30	—	—	−0.5	−1.0
3	2	11	32	1	2	+0.5	+1.0

1. Pooled OLS (Ignoring a_i): Across all 6 pooled observations, $\bar{x} = 6.67$, $\bar{y} = 20.83$, $\sum(x_{it} - \bar{x})^2 = 63.333$, and $\sum(x_{it} - \bar{x})(y_{it} - \bar{y}) = 156.667$, giving $\hat{\beta}_{POLS} = 156.667/63.333 = 2.4737$.

2. First-Differenced (FD) Estimator (No Intercept): Using the $N = 3$ differenced pairs (2, 3), (3, 4), and (1, 2) gives $\hat{\beta}_{FD} = (2 \cdot 3 + 3 \cdot 4 + 1 \cdot 2)/(2^2 + 3^2 + 1^2) = 20/14 = 1.4286$.

3. Fixed Effects (FE / Within) Estimator: Using the time-demeaned columns gives $\sum(x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i) = 10$ and $\sum(x_{it} - \bar{x}_i)^2 = 7$, so $\hat{\beta}_{FE} = 10/7 = 1.4286$. Notice two exact econometric results: first, when $T = 2$, FD and FE produce the identical slope 1.4286; second, Pooled OLS (2.4737) severely overstates the causal training effect because higher-productivity firms ($a_3 > a_2 > a_1$) also provide more training.

Two-Way Fixed Effects and Choosing Between FD and FE ($T \geq 3$)

In empirical applications across Australian states or ASX corporations over multiple years ($T \geq 3$), researchers routinely include **two-way fixed effects**: unit fixed effects a_i alongside period fixed effects λ_t (implemented via $T - 1$ year or quarter dummy variables):

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + a_i + \lambda_t + u_{it}$$

Where a_i absorbs all time-invariant unit heterogeneity and λ_t absorbs all common macroeconomic shocks affecting every unit in period t (such as RBA monetary tightening, national wage awards, or pandemic lockdowns). When $T \geq 3$, FD and FE yield different estimates: if the idiosyncratic errors u_{it}

are serially uncorrelated, FE is strictly more efficient than FD (indeed, under i.i.d. u_{it}, the first-differenced error $\Delta u_{it} = u_{it} - u_{i,t-1}$ has negative autocorrelation $\text{Corr}(\Delta u_{it}, \Delta u_{i,t-1}) = -0.5$). Conversely, if u_{it} follows a random walk ($u_{it} = u_{i,t-1} + e_{it}$), then $\Delta u_{it} = e_{it}$ is serially uncorrelated and FD is more efficient than FE. Regardless of whether FD or FE is chosen, standard errors should be clustered at the unit level i.

4.6 Limited Dependent Variables: LPM, Logit, and Probit Basics

When the outcome variable is binary ($y_i \in \{0, 1\}$)—for example, whether an Australian resident participates in the labour force (1 = employed/seeking work, 0 = not in labour force), defaults on a mortgage, or exports goods—the conditional expectation equals the conditional response probability: $E(y_i \mid \mathbf{x}_i) = P(y_i = 1 \mid \mathbf{x}_i)$.

The Linear Probability Model (LPM)

Applying OLS directly to $y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i$ defines the **Linear Probability Model (LPM)**, where β_j measures the constant change in probability $P(y_i = 1 \mid \mathbf{x}_i)$ for a one-unit increase in x_j. While straightforward to interpret, the LPM suffers from three structural flaws: (1) fitted probabilities $\hat{y}_i$ can fall outside $[0, 1]$; (2) marginal effects are constrained to be constant rather than tapering near 0 and 1; and (3) because a Bernoulli variable has conditional variance $P_i(1 - P_i)$, the error term is inherently heteroskedastic ($\text{Var}(u_i \mid \mathbf{x}_i) = P_i(1 - P_i)$), necessitating White robust standard errors.

Binary Logit and Probit Models

To guarantee predicted probabilities strictly lie in $(0, 1)$, binary response models apply a nonlinear cumulative distribution function $G(\cdot)$ to a linear index $z_i = \mathbf{x}_i'\boldsymbol{\beta}$. Equivalently, they arise from an unobserved continuous **latent variable** $y_i^* = \mathbf{x}_i'\boldsymbol{\beta} + e_i$, where we observe $y_i = 1$ if $y_i^* > 0$ and $y_i = 0$ otherwise:

$$P(y_i = 1 \mid \mathbf{x}_i) = G(\mathbf{x}_i'\boldsymbol{\beta})$$

- **Logit Model:** Assumes e_i follows a standard logistic distribution, giving the logistic CDF:

$$G(z) = \Lambda(z) = \frac{\exp(z)}{1 + \exp(z)} = \frac{1}{1 + \exp(-z)}$$

- **Probit Model:** Assumes $e_i \sim N(0, 1)$, giving the standard normal CDF:

$$G(z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{v^2}{2}\right) dv$$

DegreeLive

Because Logit and Probit are nonlinear in $\boldsymbol{\beta}$, parameters are estimated by **Maximum Likelihood Estimation (MLE)**, maximising the log-likelihood function:

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \{y_i \ln G(\mathbf{x}_i'\boldsymbol{\beta}) + (1 - y_i) \ln[1 - G(\mathbf{x}_i'\boldsymbol{\beta})]\}$$

Where $\ell(\boldsymbol{\beta})$ is the sample log-likelihood summed over all n binary observations. Importantly, raw coefficients $\hat{\beta}_j$ from Logit and Probit are *not* marginal probability effects; only their sign and statistical significance are read directly. For a continuous variable x_j, the partial effect on the response probability is:

$$\frac{\partial P(y_i = 1 \mid \mathbf{x}_i)}{\partial x_j} = g(\mathbf{x}_i'\boldsymbol{\beta}) \cdot \beta_j$$

Where $g(z) = dG(z)/dz > 0$ is the corresponding density function ($\lambda(z) = \Lambda(z)[1 - \Lambda(z)]$ for Logit; $\phi(z) = (2\pi)^{-1/2} \exp(-z^2/2)$ for Probit). Because the logistic variance is $\pi^2/3$ whereas standard normal variance is 1, Logit coefficients are roughly 1.6 to 1.8 times larger in magnitude than Probit coefficients for the same dataset, though both yield nearly identical predicted probabilities and **Average Partial Effects (APEs)**, $n^{-1} \sum_{i=1}^{n} g(\mathbf{x}_i'\hat{\boldsymbol{\beta}})\hat{\beta}_j$. In Logit models, $\exp(\hat{\beta}_j)$ also has a direct **odds-ratio** interpretation.

Worked Numerical Comparison (Australian Labour Force Participation): Using HILDA survey microdata for $n = 1,500$ prime-age adults, an analyst models labour force participation ($lfp_i \in \{0, 1\}$) on years of education (*educ*) and number of preschool children (*kids*). Evaluating at $educ = 14$ and $kids = 1$ yields the Probit linear index $\hat{z} = -0.80 + 0.12(14) - 0.35(1) = 0.53$, so $\hat{P}_{\text{Probit}} = \Phi(0.53) = 0.7019$ (70.19%). At $\hat{z} = 0.53$, the standard normal density is $\phi(0.53) = (2\pi)^{-1/2} \exp(-0.53^2/2) = 0.3467$. Therefore, the marginal effect of one extra year of schooling at this point is 0.3467 × 0.12 = +0.0416 (+4.16 percentage points), while the discrete effect of an additional preschool child (raising *kids* from 1 to 2, so $\hat{z}' = 0.18$) is $\Phi(0.18) - \Phi(0.53) = 0.5714 - 0.7019 = -0.1305$ (–13.05 percentage points). Goodness-of-fit in MLE binary models is evaluated using McFadden's pseudo-$R^2 = 1 - \ell_{UR}/\ell_0$ and the percentage correctly predicted, rather than classical R^2.

Hypothesis Testing in MLE: Likelihood Ratio (LR) Test & Model Comparison

In Logit and Probit regressions, joint hypotheses involving q restrictions ($H_0 : \beta_{k-q+1} = \cdots = \beta_k = 0$) are tested using the **Likelihood Ratio (LR) statistic** rather than an F-test, comparing the maximised log-likelihood of the unrestricted model (ℓ_{UR}) against that of the restricted model (ℓ_R):

$$LR = 2(\ell_{UR} - \ell_R) \overset{a}{\sim} \chi^2_q \quad \text{under } H_0$$

Where $\ell_{UR} \geq \ell_R$ always holds because removing restrictions can never lower the maximised log-likelihood. For example, if adding $q = 2$ state-policy indicators to a Probit mortgage-default model raises the log-likelihood from $\ell_R = -412.50$ to $\ell_{UR} = -406.80$, then $LR = 2[-406.80 - (-412.50)] = 11.40 > \chi^2_{2,0.01} = 9.21$, rejecting H_0 at the 1% level. Similarly, in Logit models, if $\hat{\beta}_{educ} = 0.20$, the odds ratio $\exp(0.20) = 1.221$ indicates that each additional year of education multiplies the odds of labour-force participation $P/(1-P)$ by 1.221 (a 22.1% increase in odds).

Feature	Linear Probability Model (LPM)	Binary Logit	Binary Probit
Link Function $G(z)$	Identity: $G(z) = z$	Logistic CDF: $\Lambda(z) = (1+e^{-z})^{-1}$	Standard Normal CDF: $\Phi(z)$
Predicted Probability Range	Unbounded (can be < 0 or > 1)	Strictly $(0,1)$	Strictly $(0,1)$
Estimation Method	OLS (with White HC robust SEs)	Maximum Likelihood (MLE)	Maximum Likelihood (MLE)
Marginal Effect of x_j	Constant: $\hat{\beta}_j$	$\Lambda(z)[1-\Lambda(z)]\hat{\beta}_j$	$\phi(\mathbf{x}'\hat{\boldsymbol{\beta}})\hat{\beta}_j$
Quick Scaling Rule vs. LPM	Baseline ($\hat{\beta}_{LPM}$)	$\hat{\beta}_{Logit} \approx 4\hat{\beta}_{LPM}$ (at $P = 0.5$)	$\hat{\beta}_{Probit} \approx 2.5\hat{\beta}_{LPM}$ (at $P = 0.5$)

Corner-Solution Outcomes: The Tobit (Censored) Model

Beyond binary 0/1 decisions, economic survey variables are frequently continuous over strictly positive values yet pile up at zero for a substantial fraction of the population—examples include annual household charitable donations, voluntary after-tax superannuation contributions, or weekly hours of paid overtime. Applying OLS to such **corner-solution** outcomes mixes the extensive margin (whether $y_i > 0$) with the intensive margin (how much y_i given $y_i > 0$) and produces biased slope estimates. The **Tobit model** specifies a latent linear variable $y_i^* = \mathbf{x}_i'\boldsymbol{\beta} + u_i$ with $u_i \mid \mathbf{x}_i \sim N(0, \sigma^2)$, where the observed outcome is $y_i = \max(0, y_i^*)$.

Estimated by Maximum Likelihood combining the probability mass at $y_i = 0$ (equal to $\Phi(-\mathbf{x}_i'\boldsymbol{\beta}/\sigma)$) and the normal density for $y_i > 0$, the Tobit model yields two distinct partial effects: the partial effect on the conditional expectation given positivity, $\partial E(y_i \mid y_i > 0, \mathbf{x}_i)/\partial x_j$, and the unconditional partial effect across the entire population, $\partial E(y_i \mid \mathbf{x}_i)/\partial x_j = \Phi(\mathbf{x}_i'\boldsymbol{\beta}/\sigma)\beta_j$. Scaling $\hat{\beta}_j$ by the probability of being uncensored $\Phi(\mathbf{x}_i'\hat{\boldsymbol{\beta}}/\hat{\sigma})$ makes Tobit marginal effects directly comparable to OLS slopes across the full sample.

DegreeLive