DegreeLive

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

Applied econometrics

Course Code: ETC3410 • Module 2 Notes

Module 2: Panel Data Models & Difference-in-Differences

2.1 Longitudinal Data Structures and the Unobserved Effects Model

Panel (or longitudinal) datasets combine cross-sectional and time-series dimensions by repeatedly observing the same economic units—workers, households, firms, local government areas, or Australian states—across multiple discrete time periods. Prominent Australian examples include the Household, Income and Labour Dynamics in Australia (HILDA) Survey tracking over 17,000 individuals annually since 2001, the Australian Bureau of Statistics (ABS) Business Longitudinal Analysis Data Environment (BLADE) linking ATO corporate tax filings across firms, and state-by-quarter labour force panels.

Let units be indexed by $i = 1, \dots, N$ and time periods by $t = 1, \dots, T$. In microeconometric panels, N is typically large (hundreds or thousands of households) while T is short and fixed (e.g., $T = 2$ to 20 waves), so asymptotic inference relies on $N \to \infty$ with T fixed. A panel is **balanced** if every unit i is observed in all T periods (total observations NT), and **unbalanced** if observation counts T_i vary across units due to entry, exit, or survey attrition. When attrition is non-random, selection bias arises unless missingness is independent of idiosyncratic shocks conditional on covariates and fixed effects.

The canonical linear **unobserved effects panel data model** decomposes the error term into a time-invariant unit-specific component and an idiosyncratic time-varying shock:

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + c_i + u_{it}, \qquad i = 1, \dots, N, \quad t = 1, \dots, T$$

Where:

- y_{it} is the scalar outcome variable for unit i in period t (e.g., log hourly wage in AUD).
- $\mathbf{x}_{it}$ is a $K \times 1$ vector of observable time-varying explanatory variables (e.g., union membership, hours worked, local unemployment rate).
- $\boldsymbol{\beta}$ is the $K \times 1$ parameter vector of primary causal interest.
- c_i is the unobserved time-invariant unit heterogeneity (e.g., innate cognitive ability, risk aversion, or persistent firm managerial quality).
- u_{it} is the idiosyncratic time-varying error term capturing shocks affecting unit i solely in period t.

2.2 Pooled OLS and Cluster-Robust Inference

If an analyst ignores the panel structure and stacks all observations into a single regression $y_{it} = \alpha + \mathbf{x}_{it}'\boldsymbol{\beta} + v_{it}$ where $v_{it} = c_i + u_{it}$ is the composite error, the resulting **Pooled OLS** estimator is consistent only if the composite error is uncorrelated with $\mathbf{x}_{it}$:

$$E[\mathbf{x}_{it}v_{it}] = E[\mathbf{x}_{it}c_i] + E[\mathbf{x}_{it}u_{it}] = \mathbf{0}$$

Where:

- $E[\mathbf{x}_{it}c_i] = \mathbf{0}$ requires unobserved heterogeneity c_i to be strictly orthogonal to all regressors. In applied economics, this condition routinely fails: high-ability workers (c_i) simultaneously attain more schooling, enter unionised jobs, and earn higher wages, imparting severe omitted variable bias to Pooled OLS.

Even when $E[\mathbf{x}_{it}v_{it}] = \mathbf{0}$ holds (for example, in a multi-wave RCT), the composite error $v_{it} = c_i + u_{it}$ exhibits positive serial correlation across time within unit i because c_i is constant across all t:

$$\text{Cov}(v_{it}, v_{is}) = \text{Var}(c_i) = \sigma_c^2 > 0 \qquad \text{for } t \neq s$$

Where:

- σ_c^2 is the variance of the time-invariant individual effect c_i.

Conventional heteroskedasticity-robust (White) standard errors ignore within-unit correlation and severely understate true sampling variability. Valid inference requires **cluster-robust (Liang-Zeger) variance estimation** clustered at the panel unit i:

$$\widehat{\text{Var}}_{\text{CR}}(\hat{\boldsymbol{\beta}}_{\text{POLS}}) = \left(\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \mathbf{X}_i'\hat{\mathbf{v}}_i\hat{\mathbf{v}}_i'\mathbf{X}_i\right)\left(\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{X}_i\right)^{-1}$$

Where:

- $\mathbf{X}_i$ is the $T_i \times K$ regressor matrix for unit i.

DegreeLive

- $\hat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}_{\text{POLS}}$ is the $T_i \times 1$ residual vector for unit i, allowing arbitrary serial correlation and heteroskedasticity within unit i while assuming independence across distinct units $i \neq j$.

2.3 Fixed Effects (Within) and First-Differenced Estimators

When c_i is arbitrarily correlated with $\mathbf{x}_{it}$—written $E[c_i \mid \mathbf{X}_i] \neq 0$—consistent estimation of $\boldsymbol{\beta}$ requires eliminating c_i via within-unit differencing transformations. Both Fixed Effects (FE) and First Differences (FD) rely on the **Strict Exogeneity** assumption:

$$E[u_{it} \mid \mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iT}, c_i] = 0 \qquad \text{for all } t = 1, \dots, T$$

Where:

- Strict exogeneity demands that idiosyncratic shocks u_{it} in period t be uncorrelated with regressors in *past, current, and future* periods ($s = 1, \dots, T$), conditional on c_i.
- Strict exogeneity is violated by lagged dependent variables ($y_{i,t-1}$ as a regressor) or feedback loops where a negative earnings shock u_{it} induces a worker to change industry or training status in $t+1$.

The Fixed Effects (Within) Estimator

Averaging the structural equation across all T periods for unit i yields the unit-mean equation $\bar{y}_i = \bar{\mathbf{x}}_i'\boldsymbol{\beta} + c_i + \bar{u}_i$. Subtracting this time average from each period observation removes c_i identically, producing the **within-transformed (time-demeaned) equation**:

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

Where:

- $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$ and $\bar{\mathbf{x}}_i = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_{it}$ are unit-specific time means.
- $\ddot{y}_{it}$ and $\ddot{\mathbf{x}}_{it}$ are time-demeaned deviations from each unit's own longitudinal average.

Applying Pooled OLS to the time-demeaned variables yields the **Fixed Effects (Within) estimator**:

$$\hat{\boldsymbol{\beta}}_{\text{FE}} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{\mathbf{x}}_{it}\ddot{\mathbf{x}}_{it}'\right)^{-1} \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{\mathbf{x}}_{it}\ddot{y}_{it}\right)$$

Where:

- Notice that any time-invariant regressor z_i (such as birth year, gender, or baseline ATAR score) satisfies $z_i - \bar{z}_i = 0$ identically; hence Fixed Effects cannot identify coefficients on time-invariant variables, though interactions between z_i and time dummies remain identifiable.

By the **Frisch-Waugh-Lovell (FWL) theorem**, $\hat{\boldsymbol{\beta}}_{\text{FE}}$ is algebraically identical to the **Least Squares Dummy Variable (LSDV)** estimator obtained by regressing y_{it} on $\mathbf{x}_{it}$ and N separate unit indicator dummies. Because estimating N implicit unit intercepts consumes N degrees of freedom, the unclustered residual variance estimator uses $N(T-1) - K$ degrees of freedom rather than $NT - K$:

$$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} \hat{\ddot{u}}_{it}^2}{N(T-1) - K}$$

Where:

- $\hat{\ddot{u}}_{it} = \ddot{y}_{it} - \ddot{\mathbf{x}}_{it}'\hat{\boldsymbol{\beta}}_{\text{FE}}$ are the within-regression residuals.
- For valid inference when u_{it} is serially correlated within units, econometricians apply cluster-robust standard errors at the unit level i (Stock & Watson, 2008).

First-Differenced (FD) Estimator vs. Fixed Effects

An alternative transformation subtracts period $t-1$ from period t for $t = 2, \dots, T$:

$$\Delta y_{it} = y_{it} - y_{i,t-1} = \Delta\mathbf{x}_{it}'\boldsymbol{\beta} + \Delta u_{it}$$

Where:

- Δy_{it} and $\Delta\mathbf{x}_{it}$ are single-period changes, eliminating c_i because $c_i - c_i = 0$.

DegreeLive

When $T = 2$, Fixed Effects and First Differences produce identical point estimates and standard errors ($hat\boldsymbol{\beta}_{\text{FE}} = \hat{\boldsymbol{\beta}}_{\text{FD}}$). When $T \geq 3$, their relative efficiency depends on the serial correlation of u_{it}: if u_{it} is serially uncorrelated, FE is more efficient (BLUE); if u_{it} follows a random walk ($u_{it} = u_{i,t-1} + e_{it}$), then $\Delta u_{it} = e_{it}$ is white noise and FD is more efficient. Moreover, if FE and FD yield substantially different point estimates when $T \geq 3$, it signals a violation of strict exogeneity or measurement error attenuation (which affects FD more severely than FE when regressors are persistent).

Two-Way Fixed Effects (TWFE)

Applied microeconometrics routinely includes both unit fixed effects c_i and common time period fixed effects λ_t to absorb aggregate macroeconomic shocks (such as RBA cash rate cycles, COVID-19 lockdowns, or national CPI inflation):

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

Where:

- c_i absorbs all time-invariant unit differences.
- λ_t absorbs all nation-wide shocks common to all units in period t, leaving variation solely from unit-specific deviations over time relative to aggregate period averages.

2.4 Random Effects (RE) GLS Estimation and Specification Tests

If unobserved heterogeneity c_i is strictly uncorrelated with all explanatory variables across every period—$E[c_i \mid \mathbf{X}_i] = 0$ (alongside $E[u_{it} \mid \mathbf{X}_i, c_i] = 0$) —then Fixed Effects remains consistent but discards all between-unit variation ($\bar{\mathbf{x}}_i$), yielding unnecessarily large standard errors and making time-invariant regressors unidentifiable. Under this orthogonality condition, the **Random Effects (RE)** Generalised Least Squares (GLS) estimator achieves full asymptotic efficiency.

Assuming homoskedastic errors $\text{Var}(c_i \mid \mathbf{X}_i) = \sigma_c^2$ and $\text{Var}(u_{it} \mid \mathbf{X}_i, c_i) = \sigma_u^2$ with serially uncorrelated u_{it}, the composite error $v_{it} = c_i + u_{it}$ has constant variance $\sigma_c^2 + \sigma_u^2$ and constant equicorrelated intraclass correlation across any two periods $t \neq s$:

$$\rho = \text{Corr}(v_{it}, v_{is}) = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_u^2}$$

Where:

- $\rho \in [0, 1)$ is the proportion of composite error variance attributable to permanent unit-specific heterogeneity c_i.

The exact GLS transformation **quasi-demeans** each variable by subtracting a fraction $\theta \in [0, 1)$ of its unit time mean:

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

Where:

- θ is the quasi-demeaning weight. When $\sigma_c^2 = 0, \theta = 0$ and RE collapses to Pooled OLS. As $T \to \infty$ or $\sigma_c^2/\sigma_u^2 \to \infty, \theta \to 1$ and RE converges to Fixed Effects.
- Feasible GLS (FGLS) replaces σ_u^2 and σ_c^2 with consistent estimates $\hat{\sigma}_u^2$ (from FE residuals) and $\hat{\sigma}_c^2$ (from between-effects or Pooled OLS residuals).

Breusch-Pagan Lagrange Multiplier (LM) Test

To test whether Pooled OLS suffices against Random Effects ($H_0 : \sigma_c^2 = 0$ vs. $H_1 : \sigma_c^2 > 0$), Breusch and Pagan (1980) derived a Lagrange Multiplier test using only Pooled OLS residuals $\hat{v}_{it}$:

$$LM = \frac{NT}{2(T-1)}\left[\frac{\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\hat{v}_{it}\right)^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{v}_{it}^2} - 1\right]^2 \overset{d}{\sim} \chi_1^2$$

Where:

- Under $H_0 : \sigma_c^2 = 0$, LM follows an asymptotic χ_1^2 distribution (or half-chi_1^2 accounting for the one-sided boundary $\sigma_c^2 \geq 0$). Rejecting H_0 indicates significant unobserved unit heterogeneity, ruling out unadjusted Pooled OLS inference.

Degree Live

Hausman Specification Test and Mundlak Correlated Random Effects

To decide between Fixed Effects and Random Effects, econometricians test the key orthogonality assumption $H_0 : \text{Cov}(c_i, \mathbf{x}_{it}) = \mathbf{0}$ using the **Hausman (1978) specification test**. Under H_0, both $\hat{\boldsymbol{\beta}}_{\text{FE}}$ and $\hat{\boldsymbol{\beta}}_{\text{RE}}$ are consistent, with $\hat{\boldsymbol{\beta}}_{\text{RE}}$ asymptotically efficient. Under the alternative $H_1 : \text{Cov}(c_i, \mathbf{x}_{it}) \neq \mathbf{0}$, $\hat{\boldsymbol{\beta}}_{\text{FE}}$ remains consistent while $\hat{\boldsymbol{\beta}}_{\text{RE}}$ is biased and inconsistent. Exploiting the zero asymptotic covariance between an efficient estimator and the difference $(\hat{\boldsymbol{\beta}}_{\text{FE}} - \hat{\boldsymbol{\beta}}_{\text{RE}})$ under H_0, the Hausman statistic is:

$$H = (\hat{\boldsymbol{\beta}}_{\text{FE}} - \hat{\boldsymbol{\beta}}_{\text{RE}})' \left[\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_{\text{FE}}) - \widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_{\text{RE}})\right]^{-1} (\hat{\boldsymbol{\beta}}_{\text{FE}} - \hat{\boldsymbol{\beta}}_{\text{RE}}) \xrightarrow{d} \chi^2_K$$

Where:

- K is the number of time-varying regressors compared between the FE and RE models.
- Rejection of H_0 ($H > \chi^2_{K,1-\alpha}$) implies c_i is correlated with $\mathbf{x}_{it}$, requiring Fixed Effects rather than Random Effects.

A practical limitation of the classical Hausman test is that it requires homoskedasticity and no serial correlation in u_{it} so that $\hat{\boldsymbol{\beta}}_{\text{RE}}$ is fully efficient; otherwise $\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_{\text{FE}}) - \widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_{\text{RE}})$ can fail to be positive definite. The modern robust alternative is the **Mundlak (1978) Correlated Random Effects (CRE)** regression, which models $c_i = \mathbf{x}\text{-mean}_i'\boldsymbol{\gamma} + a_i = \bar{\mathbf{x}}_i'\boldsymbol{\gamma} + a_i$ where $a_i \perp \mathbf{X}_i$:

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

Where:

- $\bar{\mathbf{x}}_i$ is the vector of unit-level time averages for all time-varying regressors.
- Estimating this augmented equation by RE-GLS produces $\hat{\boldsymbol{\beta}}_{\text{CRE}} \equiv \hat{\boldsymbol{\beta}}_{\text{FE}}$ identically, while permitting time-invariant controls z_i to remain in the regression.
- Testing the joint Wald hypothesis $H_0 : \boldsymbol{\gamma} = \mathbf{0}$ using **cluster-robust standard errors** provides an asymptotically valid, positive-definite, cluster-robust Hausman test.

Worked Numerical Example 2.1: Quasi-Demeaning Parameter θ and Hausman Test

An applied econometrician uses $T = 5$ annual waves of HILDA data ($N = 1,200$ workers) to estimate the wage return to union membership ($x_{it} \in \{0, 1\}$) via log hourly wages y_{it}. FE estimation produces $\hat{\beta}_{\text{FE}} = 0.068$ ($\text{SE}_{\text{FE}} = 0.015$) and residual variance $\hat{\sigma}^2_u = 0.036$. Random Effects estimation yields $\hat{\beta}_{\text{RE}} = 0.095$ ($\text{SE}_{\text{RE}} = 0.011$) and unobserved heterogeneity variance $\hat{\sigma}^2_c = 0.054$.

Step 1 — Compute the Intraclass Correlation $\hat{\rho}$ and Quasi-Demeaning Weight $\hat{\theta}$:

$$\hat{\rho} = \frac{\hat{\sigma}^2_c}{\hat{\sigma}^2_c + \hat{\sigma}^2_u} = \frac{0.054}{0.054 + 0.036} = \frac{0.054}{0.090} = 0.60$$

$$\hat{\theta} = 1 - \sqrt{\frac{0.036}{0.036 + 5(0.054)}} = 1 - \sqrt{\frac{0.036}{0.306}} = 1 - \sqrt{0.11765} \approx 1 - 0.3430 = 0.6570$$

Where:

- 60% of the composite wage error variance reflects permanent individual heterogeneity c_i, and RE subtracts 65.7% of each worker's 5-year mean wage and union status.

Step 2 — Compute the Hausman Specification Statistic ($K = 1$):

$$H = \frac{(\hat{\beta}_{\text{FE}} - \hat{\beta}_{\text{RE}})^2}{\text{SE}^2_{\text{FE}} - \text{SE}^2_{\text{RE}}} = \frac{(0.068 - 0.095)^2}{0.015^2 - 0.011^2} = \frac{(-0.027)^2}{0.000225 - 0.000121} = \frac{0.000729}{0.000104} \approx 7.01$$

Where:

- Comparing $H = 7.01$ against the $\chi^2_{1,0.95} = 3.84$ (and $\chi^2_{1,0.99} = 6.63$) critical values rejects $H_0 : \text{Cov}(c_i, x_{it}) = 0$ at the 1% level.
- Conclusion: Union membership is correlated with unobserved worker productivity c_i, so RE is inconsistent and the FE estimate (6.8% wage premium) must be retained.

DegreeLive

2.5 Canonical Difference-in-Differences (DiD)

When a policy reform occurs at a sharp calendar date for one jurisdiction or group while another group remains unexposed—for example, Victoria altering state payroll tax thresholds in 2019 while New South Wales keeps its existing schedule—econometricians employ **Difference-in-Differences (DiD)**. Unlike simple cross-sectional comparisons (biased by permanent state differences) or before-after comparisons (biased by macroeconomic trends), DiD combines both contrasts.

In the canonical 2×2 setting, there are two groups $G_i \in \{0,1\}$ (treated group $G_i = 1$, control group $G_i = 0$) and two periods $t \in \{0,1\}$ (pre-treatment $t = 0$, post-treatment $t = 1$). Treatment activates solely for group $G_i = 1$ in period $t = 1$, so observed treatment status is $D_{it} = G_i \cdot \mathbf{1}(t = 1)$. The target estimand is the **Average Treatment Effect on the Treated in the post-period**:

$$\tau_{\text{ATT}} = E[Y_{i1}(1) - Y_{i1}(0) \mid G_i = 1]$$

Where:

- $Y_{i1}(1)$ and $Y_{i1}(0)$ are unit i's treated and untreated potential outcomes in period $t = 1$.

Identifying Assumptions: Parallel Trends and No Anticipation

Identification of τ_{ATT} rests on two foundational assumptions:

1. **No Anticipation:** Prior to policy implementation ($t = 0$), future treatment status has zero causal effect on outcomes: $Y_{i0}(1) = Y_{i0}(0) = Y_{i0}$ for all treated units ($G_i = 1$). If firms delay hiring or pull forward capital expenditure in $t = 0$ anticipating a tax change in $t = 1$, baseline $t = 0$ is contaminated.
2. **Parallel Trends Assumption:** In the absence of treatment, the treated group would have experienced the exact same average outcome trajectory from $t = 0$ to $t = 1$ as the control group:

$$E[Y_{i1}(0) - Y_{i0}(0) \mid G_i = 1] = E[Y_{i1}(0) - Y_{i0}(0) \mid G_i = 0]$$

Where:

- $E[Y_{i1}(0) - Y_{i0}(0) \mid G_i = 1]$ is the unobservable counterfactual trend of the treated group.
- Notice that Parallel Trends allows treated and control groups to have completely different baseline outcome *levels* ($E[Y_{i0}(0) \mid G_i = 1] \neq E[Y_{i0}(0) \mid G_i = 0]$), provided their *changes* over time would be parallel without treatment. Crucially, Parallel Trends is **functional-form dependent**: if it holds in levels (Y), it generally cannot hold simultaneously in logarithms ($\ln Y$) unless baseline levels are equal.

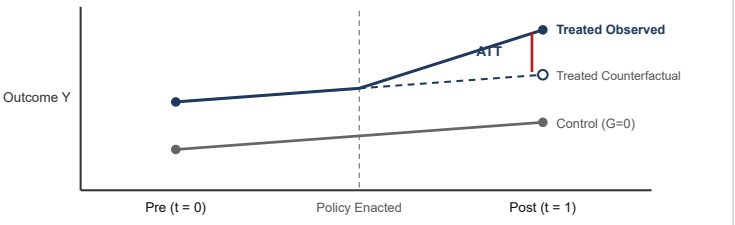
Under No Anticipation and Parallel Trends, the population DiD estimand exactly recovers τ_{ATT}:

$$\tau_{\text{DiD}} = (E[Y_{i1} \mid G_i = 1] - E[Y_{i0} \mid G_i = 1]) - (E[Y_{i1} \mid G_i = 0] - E[Y_{i0} \mid G_i = 0]) = \tau_{\text{ATT}}$$

Where:

- The first bracket measures the before-after change for the treated group ($\tau_{\text{ATT}} + \text{time trend}$), and the second bracket measures the before-after change for the control group (time trend).

Figure 2.1: Canonical 2x2 Difference-in-Differences & Parallel Trends



Regression Implementation and Cluster Inference

In the canonical 2×2 framework, the DiD estimator is equivalently obtained by OLS in an interaction regression:

$$Y_{it} = \beta_0 + \beta_1 G_i + \beta_2 \text{Post}_t + \tau_{\text{DiD}}(G_i \times \text{Post}_t) + u_{it}$$

Where:

DegreeLive

- $\beta_0 = E[Y_{i0} \mid G_i = 0]$ is the baseline mean of the control group.
- β_1 captures the permanent pre-existing gap between treated and control groups.
- β_2 captures the common macro time trend shared by both groups.
- $\hat{\tau}_{\text{DiD}}$ is numerically identical to $(\bar{Y}_{1,1} - \bar{Y}_{1,0}) - (\bar{Y}_{0,1} - \bar{Y}_{0,0})$ and equals the TWFE regression $Y_{it} = c_i + \lambda_t + \tau_{\text{DiD}} D_{it} + u_{it}$ in balanced 2×2 panels.

As Bertrand, Duflo, and Mullainathan (2004) demonstrated, outcomes in multi-year policy panels (such as state unemployment or firm employment) exhibit strong positive serial correlation. Standard OLS or unclustered White errors dramatically over-reject true nulls. Inference requires **clustering standard errors at the unit/jurisdiction of treatment assignment** (e.g., state or LGA). When the number of treated or total clusters is small (fewer than 30–40 clusters, typical across Australia's 8 states and territories), cluster-robust asymptotics fail, and applied researchers implement the **Wild Cluster Bootstrap** (Cameron, Gelbach, & Miller, 2008) or permutation-based randomisation inference.

2.6 Event Studies and Dynamic Pre-Trends Testing

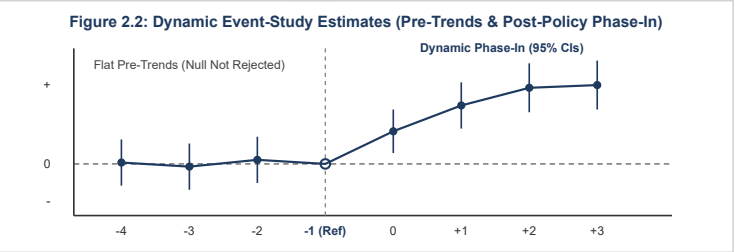
When panel data span multiple pre-treatment and post-treatment periods ($t = -L, \dots, -1, 0, 1, \dots, M$ relative to treatment onset at $t = 0$), researchers generalise the static 2×2 model into a dynamic **Event Study** specification:

$$Y_{it} = c_i + \lambda_t + \sum_{k=-L,\, k \neq -1}^{M} \delta_k \cdot \mathbf{1}(t - E_i = k) + u_{it}$$

Where:

- E_i is the calendar period in which unit i first receives treatment (so $t - E_i = k$ measures event time relative to reform).
- $k = -1$ (the period immediately prior to reform) is omitted as the reference baseline, normalising $\delta_{-1} = 0$.
- Lead coefficients ($k \leq -2$) test for **pre-trends**: statistically indistinguishable estimates near zero ($\delta_{-L} \approx \cdots \approx \delta_{-2} \approx 0$) provide supporting falsification evidence consistent with parallel pre-treatment trajectories.
- Lag coefficients ($k \geq 0$) trace out the dynamic trajectory of the treatment effect—revealing immediate impacts, gradual phase-in, or long-run fade-out.

Econometricians must treat pre-trend tests carefully: Roth (2022) showed that failing to reject $H_0 : \delta_{-2} = 0$ in underpowered panels does not prove parallel trends hold (low power against linear violations), and conditioning publication or post-period analysis on passing a pre-trend test introduces pre-test selection bias. Complementary sensitivity analysis (Rambachan & Roth, 2023, "Honest DiD") bounds how large post-treatment violations of parallel trends could be while still preserving a statistically meaningful ATT.



2.7 Staggered Adoption and Modern DiD Diagnostics

Many major policy reforms in Australia and internationally roll out across states, municipalities, or firms at different points in calendar time—known as **staggered treatment adoption** (for instance, progressive state-by-state adoption of container deposit schemes or digital payroll reporting). Historically, researchers estimated a single static Two-Way Fixed Effects (TWFE) model:

$$Y_{it} = c_i + \lambda_t + \beta^{\text{TWFE}} D_{it} + u_{it}$$

Where:

- $D_{it} = 1$ if unit i has already adopted the policy by period t, and 0 otherwise.

The Goodman-Bacon (2021) Decomposition

Goodman-Bacon (2021) proved that the static $\hat{\beta}^{\text{TWFE}}$ estimator is a variance-weighted average of *all* possible 2×2 DiD comparisons in the dataset. With three timing groups—Early adopters (E), Late adopters (L), and Never-treated units (U)—TWFE decomposes into four types of 2×2 contrasts:

1. **Early vs. Never-treated (E vs. U):** Clean comparison using untreated units as controls.
2. **Late vs. Never-treated (L vs. U):** Clean comparison using untreated units as controls.
3. **Early vs. Late (before L is treated):** Clean comparison using not-yet-treated units as controls.

4. **Late vs. Early (after E is already treated): Forbidden comparison!** Already-treated Early adopters (E) act as the "control group" for Late adopters (L).

If treatment effects evolve dynamically over time after adoption (e.g., growing larger over years 1–4), subtracting the changing post-treatment trajectory of Early adopters (E) biases the Late-vs-Early 2×2 contrast downward—and can even flip the sign of $\hat{\beta}^{\text{TWFE}}$ to negative when every unit's true causal effect is strictly positive! Moreover, de Chaisemartin and D'Haultfoeuille (2020) showed that TWFE weights on certain cohort-time cells can become strictly negative.

Heterogeneity-Robust Staggered DiD Estimators

To eliminate forbidden comparisons and negative weights, modern applied econometrics relies on heterogeneity-robust estimators:

Estimator	Core Identification Idea	Control Group Used	Key Advantage
Callaway & Sant'Anna (2021)	Estimates disaggregated cohort-time $\text{ATT}(g,t)$ for every cohort g and period t, then aggregates with non-negative weights	Never-treated ($G = \infty$) or Not-yet-treated ($G > t$) units	Supports outcome regression, IPW, and doubly robust conditional parallel trends
Sun & Abraham (2021)	Interacts cohort indicators with event-time dummies in a fully saturated regression, weighting by cohort sample shares	Last-treated or Never-treated cohort	Directly yields clean dynamic event-study plots free of cross-lag contamination
Borusyak, Jaravel & Spiess (2024)	Imputation estimator: fits $Y_{it}(0) = c_i + \lambda_t + u_{it}$ solely on untreated (i,t) observations, then computes $\hat{\tau}_{it} = Y_{it} - \hat{Y}_{it}(0)$	All untreated (i,t) observations jointly	Maximum efficiency under spherical errors across long pre-periods

Worked Numerical Example 2.2: Canonical DiD & Staggered TWFE Forbidden Comparison Bias

Consider an evaluation of an Australian regional small-business digitalisation subsidy rolled out across three equally sized LGA groups over $T = 3$ years ($t = 1, 2, 3$):

- **Group E (Early Adopter):** Treated in $t = 2$ and $t = 3$. Baseline productivity index is 100 in $t = 1$; counterfactual untreated trajectory grows by $+2$ per year $(100, 102, 104)$. Treatment delivers a dynamic effect of $+10$ in its first active year ($t = 2$) and $+24$ in its second active year ($t = 3$). Observed outcomes: $(100, 112, 128)$.
- **Group L (Late Adopter):** Treated only in $t = 3$. Counterfactual untreated trajectory grows identically by $+2$ per year from baseline 90: $(90, 92, 94)$. Treatment delivers $+10$ in its first active year ($t = 3$). Observed outcomes: $(90, 92, 104)$.
- **Group U (Never-Treated Control):** Never treated. Observed outcomes from baseline 80: $(80, 82, 84)$.

Step 1 — Verify True Cohort-Time ATTs (Callaway & Sant'Anna using U as control):

$$\text{ATT}(E,2) = (112 - 100) - (82 - 80) = 12 - 2 = +10$$

$$\text{ATT}(E,3) = (128 - 100) - (84 - 80) = 28 - 4 = +24$$

$$\text{ATT}(L,3) = (104 - 92) - (84 - 82) = 12 - 2 = +10$$

Where:

- Every treated observation experiences a strictly positive causal gain ($+10$, $+24$, $+10$), with simple average ATT across the three treated cell-years equal to $(10 + 24 + 10)/3 = +14.67$.

Step 2 — Inspect the Forbidden 2×2 Comparison (L vs. Already-Treated E from $t = 2$ to $t = 3$):

$$\underline{\text{DiD}}_{L \text{ vs } E,\ (2\to3)} = (Y_{L,3} - Y_{L,2}) - (Y_{E,3} - Y_{E,2}) = (104 - 92) - (128 - 112) = 12 - 16 = -4$$

Where:

- Although Late adopters genuinely gained $+10$ in $t = 3$, using Early adopters as the "control" subtracts their $+14$ treatment effect acceleration ($+24 - +10$) on top of the $+2$ trend, producing a spurious **negative** estimate of -4!
- In a symmetric 3-group, 3-period TWFE regression without never-treated group U (comparing only E and L), TWFE averages the clean early contrast $\underline{\text{DiD}}_{E \text{ vs } L,\ (1\to2)} = (112 - 100) - (92 - 90) = +10$ and the forbidden late contrast -4, severely understating the true average gain ($+10$ and $+10$ in year 1 of adoption) and yielding $\hat{\beta}^{\text{TWFE}} = 3.0$.

DegreeLive

2.8 Australian Empirical Policy Applications

Panel data and DiD methods underpin empirical policy evaluation across Australian federal and state institutions:

- **Victorian Payroll Tax and SME Employment (BLADE Administrative Panel):** Researchers link monthly Single Touch Payroll (STP) records across Victorian and NSW firms around regional payroll-tax rate cuts (e.g., Victoria's lower regional rate introduced in 2017–2019). Using NSW border LGAs as controls and firm fixed effects c_i plus industry-by-time fixed effects λ_{st}, event studies test pre-trends in headcount before estimating employment elasticities.
- **JobKeeper Wage Subsidy Evaluation (2020–2021):** Firms experiencing a 30% revenue drop relative to the prior year qualified for JobKeeper (A$1,500 per fortnight initially). Comparing firms just above versus just below eligibility thresholds over monthly STP panels (combining DiD with threshold controls) identifies employment retention effects while controlling for sectoral COVID-19 shocks via λ_{st}.
- **Fair Work Commission Penalty Rate Reform (2017):** The reduction in Sunday penalty rates in retail and hospitality affected award-reliant employers relative to enterprise-agreement or salaried sectors. Panel regressions using HILDA and ABS Survey of Employee Earnings and Hours separate secular weekend-work trends from causal hours-worked responses.
- **Synthetic Control Method (Abadie, Diamond & Hainmueller, 2010):** When only a single Australian state implements a unique macro policy (such as Western Australia's domestic gas reservation policy or Queensland's coal royalty tier adjustment), a single unweighted control state may fail parallel trends. Synthetic Control constructs a convex combination $\sum_{j \neq 1} w_j Y_{jt}$ ($w_j \geq 0, \sum w_j = 1$) of donor states matching the treated state's pre-policy outcome trajectory and predictors, and conducts inference via in-space placebo permutations across donor states.

2.9 Triple Differences (DDD) and Covariate-Adjusted Parallel Trends

In many Australian policy settings, simple two-group parallel trends between two states may be compromised by state-specific macroeconomic shocks (such as a mining boom in Western Australia or divergent COVID-19 border closures in Victoria versus New South Wales). When an intervention targets only a specific demographic or industry subgroup *within* the treated state—for example, a Victorian payroll tax exemption applying solely to regional hospitality businesses ($H_i = 1$) while leaving non-hospitality firms ($H_i = 0$) untouched—researchers implement a **Difference-in-Difference-in-Differences (Triple Differences, DDD)** design.

Let $S_i \in \{0, 1\}$ indicate the treated jurisdiction, $H_i \in \{0, 1\}$ indicate the eligible subgroup, and $\text{Post}_t \in \{0, 1\}$ indicate the post-reform period. The DDD estimator takes the difference between the within-state DiD for the eligible subgroup and the within-state DiD for the ineligible subgroup:

$$\hat{\tau}_{\text{DDD}} = [(\bar{Y}_{S=1,H=1,\text{Post}} - \bar{Y}_{S=1,H=1,\text{Pre}}) - (\bar{Y}_{S=0,H=1,\text{Post}} - \bar{Y}_{S=0,H=1,\text{Pre}})] - [(\bar{Y}_{S=1,H=0,\text{Post}} - \bar{Y}_{S=1,H=0,\text{Pre}}) - (\bar{Y}_{S=0,H=0,\text{Post}} - \bar{Y}_{S=0,H=0,\text{Pre}})]$$

Where:

- The first bracket removes national subgroup-specific shocks affecting eligible units everywhere ($H = 1$).
- The second bracket measures state-specific macro shocks affecting ineligible units ($H = 0$) in the treated state relative to the control state.
- Subtracting the second bracket from the first cancels both state-wide macroeconomic shocks and national industry-wide trends, requiring only a weaker *relative parallel trends* assumption.

In regression form, DDD includes all main effects, pairwise two-way interactions, and the triple interaction $S_i \cdot H_i \cdot \text{Post}_t$, or equivalently state-by-time fixed effects λ_{st}, subgroup-by-time fixed effects γ_{ht}, and unit/state-by-subgroup fixed effects α_{sh}:

$$Y_{ish t} = \alpha_{sh} + \lambda_{st} + \gamma_{ht} + \tau_{\text{DDD}}(S_i \cdot H_i \cdot \text{Post}_t) + u_{ish t}$$

Where:

- α_{sh} absorbs permanent state-by-subgroup baseline differences.
- λ_{st} absorbs arbitrary state-specific time shocks (such as state lockdowns or state budget cycles).
- γ_{ht} absorbs arbitrary national subgroup-specific trends (such as national hospitality demand shifts).

Worked Numerical Example 2.3: Triple-Difference Evaluation of State Stamp Duty Reform

Suppose New South Wales ($S = 1$) eliminates stamp duty for first-home buyers ($H = 1$) purchasing homes under A$800,000 in period $\text{Post} = 1$, while repeat buyers ($H = 0$) remain subject to stamp duty. Victoria ($S = 0$) leaves its stamp duty schedule unchanged across both buyer types. Observed quarterly home-purchase rates (transactions per $1,000$ households) are recorded below:

State & Buyer Group	Pre-Reform ($t = 0$)	Post-Reform ($t = 1$)	Before–After Change (Δ)
NSW First-Home Buyers ($S = 1, H = 1$)	14.2	19.8	+5.6
Victoria First-Home Buyers ($S = 0, H = 1$)	13.5	15.1	+1.6
NSW Repeat Buyers ($S = 1, H = 0$)	22.0	24.5	+2.5
Victoria Repeat Buyers ($S = 0, H = 0$)	21.4	22.7	+1.3

Step 1 — Compute the Naive Two-State DiD for First-Home Buyers Only:

$$\widehat{\text{DiD}}_{H=1} = (19.8 - 14.2) - (15.1 - 13.5) = 5.6 - 1.6 = +4.0$$

Where:

- This $+4.0$ estimate conflates the stamp duty exemption with any NSW-wide housing credit boom between $t = 0$ and $t = 1$.

Step 2 — Compute the Placebo DiD for Ineligible Repeat Buyers ($H = 0$):

$$\widehat{\text{DiD}}_{H=0} = (24.5 - 22.0) - (22.7 - 21.4) = 2.5 - 1.3 = +1.2$$

Where:

- Repeat buyers in NSW grew faster than in Victoria by $+1.2$ transactions per $1,000$ households despite being ineligible for the exemption, revealing an underlying NSW-specific housing market expansion.

Step 3 — Compute the Triple-Difference (DDD) Estimate:

$$\hat{\tau}_{\text{DDD}} = \widehat{\text{DiD}}_{H=1} - \widehat{\text{DiD}}_{H=0} = 4.0 - 1.2 = +2.8$$

Where:

- After netting out the $+1.2$ state-wide NSW housing boom, the stamp duty exemption caused an increase of 2.8 transactions per $1,000$ eligible first-home-buyer households.

2.10 Conditional Parallel Trends and Doubly Robust DiD

In observational policy evaluations, unconditional parallel trends often fail because treated and control units have different baseline covariates X_i (for example, age, firm size, or pre-reform industry mix) that generate divergent outcome trends during macroeconomic shocks. Replacing unconditional parallel trends with **Conditional Parallel Trends** requires:

$$E[Y_{i1}(0) - Y_{i0}(0) \mid X_i, G_i = 1] = E[Y_{i1}(0) - Y_{i0}(0) \mid X_i, G_i = 0]$$

Where:

- Units sharing identical pre-treatment covariates X_i would follow parallel counterfactual paths absent treatment, alongside overlap $P(G_i = 1 \mid X_i) < 1 - \epsilon$.

Crucially, simply inserting time-varying or baseline covariates X_i additively into a TWFE regression fails to recover ATT when treatment effects vary with X_i. Instead, Sant'Anna and Zhao (2020) derive the **Doubly Robust DiD (DR-DiD)** estimator combining outcome regression on $\Delta Y_i = Y_{i1} - Y_{i0}$ among controls with inverse propensity score reweighting $p(X_i) = P(G_i = 1 \mid X_i)$:

$$\hat{\tau}_{\text{DR-DiD}} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{G_i}{\bar{G}} - \frac{\frac{\hat{p}(X_i)(1-G_i)}{1-\hat{p}(X_i)}}{E_N\left[\frac{\hat{p}(X)(1-G)}{1-\hat{p}(X)}\right]} \right) (\Delta Y_i - \hat{m}_0(X_i))$$

Where:

- $\hat{m}_0(X_i) = E[\Delta Y_i \mid X_i, G_i = 0]$ is the fitted outcome-change model estimated solely on control units.
- $\hat{p}(X_i)$ is the estimated probability of belonging to the treated group given baseline covariates X_i.
- $\hat{\tau}_{\text{DR-DiD}}$ is consistent for ATT whenever *either* the outcome evolution model $m_0(X_i)$ *or* the propensity score $p(X_i)$ is correctly specified.

2.11 Dynamic Panel Models, Nickell Bias, and Arellano-Bond GMM

Economic behaviour frequently exhibits state dependence—past household consumption, firm R&D investment, or unemployment spells directly influence current outcomes. Adding a lagged dependent variable $y_{i,t-1}$ alongside unobserved heterogeneity c_i yields the **dynamic panel model**:

$$y_{it} = \gamma y_{i,t-1} + \mathbf{x}_{it}'\boldsymbol{\beta} + c_i + u_{it}, \qquad |\gamma| < 1$$

DegreeLive

Where:

- γ is the persistence (state dependence) parameter.
- u_{it} is assumed serially uncorrelated conditional on past outcomes and covariates.

Applying the standard Fixed Effects (within) transformation fails in short panels because the demeaned lagged outcome $\ddot{y}_{i,t-1} = y_{i,t-1} - \bar{y}_{i,-1}$ correlates mechanically with the demeaned error $\ddot{u}_{it} = u_{it} - \bar{u}_i$ (specifically, $u_{i,t-1}$ inside $\bar{u}_i$ correlates positively with $y_{i,t-1}$, while u_{it} inside $\bar{u}_i$ correlates with future terms). Nickell (1981) proved that for fixed T and $N \to \infty$, the FE estimator of γ suffers a downward asymptotic bias of order $O(1/T)$:

$$\text{plim}_{N\to\infty}(\hat{\gamma}_{\text{FE}} - \gamma) \approx -\frac{1+\gamma}{T-1}$$

Where:

- For $T = 4$ and true $\gamma = 0.60$, Nickell bias is approximately $-(1.60)/3 \approx -0.53$, wiping out almost the entire persistence coefficient.
- Conversely, Pooled OLS omits c_i, imparting upward bias on $\hat{\gamma}_{\text{POLS}}$ because $y_{i,t-1}$ correlates positively with c_i. Thus, consistent estimates of γ typically lie strictly between $\hat{\gamma}_{\text{FE}}$ (lower bound) and $\hat{\gamma}_{\text{POLS}}$ (upper bound).

Arellano-Bond Difference GMM Estimator

To eliminate c_i without introducing Nickell correlation, first-differencing yields:

$$\Delta y_{it} = \gamma \Delta y_{i,t-1} + \Delta \mathbf{x}_{it}'\boldsymbol{\beta} + \Delta u_{it}, \qquad t = 3, \dots, T$$

Where:

- $\Delta y_{i,t-1} = y_{i,t-1} - y_{i,t-2}$ remains correlated with $\Delta u_{it} = u_{it} - u_{i,t-1}$ via $u_{i,t-1}$.

Anderson and Hsiao (1981) noted that deeper lags $y_{i,t-2}$ (or $\Delta y_{i,t-2}$) are correlated with $\Delta y_{i,t-1}$ yet orthogonal to $\Delta u_{it} = u_{it} - u_{i,t-1}$ provided u_{it} is serially uncorrelated. Arellano and Bond (1991) exploited *all* available past level lags $y_{i,t-2}, y_{i,t-3}, \dots, y_{i1}$ as sequential moment conditions in a Generalised Method of Moments (GMM) framework:

$$E[y_{i,t-s}\Delta u_{it}] = 0 \qquad \text{for all } s \geq 2,\ t = 3, \dots, T$$

Where:

- Valid dynamic panel GMM estimation requires passing two diagnostic tests: (1) the **Arellano-Bond AR(2) test** on differenced residuals $\Delta\hat{u}_{it}$ (which must exhibit negative first-order correlation AR(1) by construction, but *no* second-order correlation AR(2) so $y_{i,t-2}$ remains valid); and (2) the **Hansen J-test** of overidentifying restrictions (avoiding "too many instruments" relative to N by collapsing the instrument matrix).

2.12 Summary Comparison of Panel Estimators & Worked Within-Calculation

Selecting an appropriate panel estimator requires aligning identification assumptions regarding unobserved heterogeneity c_i, exogeneity of regressors $\mathbf{x}_{it}$, and serial dependence of u_{it}:

Estimator	Allows $\text{Cov}(c_i, \mathbf{x}_{it}) \neq 0$?	Identifies Time-Invariant z_i?	Exogeneity Requirement	Primary Use Case
Pooled OLS	No (inconsistent)	Yes	Contemporaneous exogeneity $E[\mathbf{x}_{it}v_{it}] = 0$	Descriptive benchmarks or multi-wave RCTs with cluster SEs
Random Effects (GLS)	No (inconsistent)	Yes	Strict exogeneity + $E[c_i \mid \mathbf{X}_i] = 0$	Efficient when c_i is orthogonal to covariates (passes Hausman test)
Fixed Effects (Within)	Yes	No (absorbed into c_i)	Strict exogeneity $E[u_{it} \mid \mathbf{X}_i, c_i] = 0$	Observational micro-panels & canonical TWFE policy models
First Differences (FD)	Yes	No (differenced out)	Strict exogeneity (or sequential for 2-period)	Random-walk errors Δu_{it} or diagnostic check against FE
Mundlak CRE	Yes	Yes	Strict exogeneity + linear $E[c_i \mid \bar{\mathbf{x}}_i]$	Cluster-robust Hausman test & non-linear panel models
Arellano-Bond GMM	Yes	No (differenced out)	Sequential exogeneity (predetermined $y_{i,t-1}$)	Short-T dynamic panels with lagged dependent variables

DegreeLive

Worked Numerical Example 2.4: Manual Within-Transformation vs. Pooled OLS

Suppose an analyst tracks $N = 3$ Victorian manufacturing SMEs over $T = 3$ years (9 firm-year observations) to estimate how annual workplace training expenditure per employee (x_{it}, in A$1,000s) affects output per worker (y_{it}, in A$1,000s):

Firm i	Year t	x_{it}	y_{it}	Firm Mean $\bar{x}_i$	Firm Mean $\bar{y}_i$	$\ddot{x}_{it} = x_{it} - \bar{x}_i$	$\ddot{y}_{it} = y_{it} - \bar{y}_i$
Firm 1	1, 2, 3	$1, 2, 3$	$50, 52, 54$	2	52	$-1, 0, +1$	$-2, 0, +2$
Firm 2	1, 2, 3	$3, 4, 5$	$68, 70, 72$	4	70	$-1, 0, +1$	$-2, 0, +2$
Firm 3	1, 2, 3	$5, 6, 7$	$86, 88, 90$	6	88	$-1, 0, +1$	$-2, 0, +2$

Step 1 — Compute the Fixed Effects (Within) Slope $\hat{\beta}_{\text{FE}}$:

$$\hat{\beta}_{\text{FE}} = \frac{\sum_{i=1}^{3}\sum_{t=1}^{3} \ddot{x}_{it}\ddot{y}_{it}}{\sum_{i=1}^{3}\sum_{t=1}^{3} \ddot{x}_{it}^2} = \frac{3 \cdot [(-1)(-2) + 0 + (1)(2)]}{3 \cdot [(-1)^2 + 0 + 1^2]} = \frac{12}{6} = 2.00$$

Where:

- Within every individual firm, each additional A$1,000 of training raises worker productivity by exactly A$2,000 ($\hat{\beta}_{\text{FE}} = 2.00$).

Step 2 — Contrast Against Pooled OLS Bias: Across all 9 observations, overall means are $\bar{x} = 4$ and $\bar{y} = 70$. Note that Firm 3 has permanently higher baseline productivity ($c_3 = 76$) and higher training ($\bar{x}_3 = 6$) than Firm 1 ($c_1 = 48, \bar{x}_1 = 2$). Computing $\sum(x_{it} - 4)^2 = 30$ and $\sum(x_{it} - 4)(y_{it} - 70) = (-3)(-20) + (-2)(-18) + (-1)(-16) + (0)(0) + \cdots = 228$ yields:

$$\hat{\beta}_{\text{POLS}} = \frac{228}{30} = 7.60$$

Where:

- Because high-productivity firms (c_i) simultaneously spend more on training ($\text{Cov}(c_i, x_{it}) > 0$), Pooled OLS ($7.60$) overstates the true causal training return (2.00) by nearly fourfold! Fixed Effects completely eliminates this cross-firm unobserved heterogeneity bias.

DegreeLive