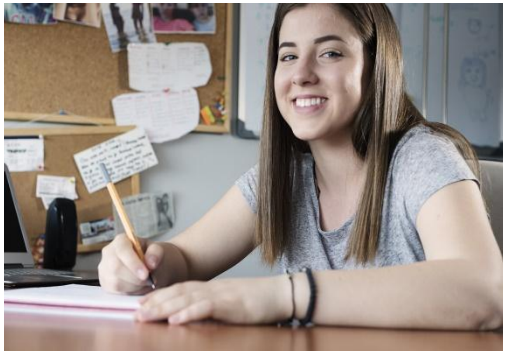


Growing Up in Ireland
National Longitudinal Study of Children

Time, Change and Causality
or
Getting the Most from GUI

Richard Layte

ESRI

Trinity College Dublin
The University of Dublin

An Roinn Leanaí agus Gnóthaí Óige
Department of Children and Youth Affairs

Introduction

- Time, change and causality
  - When is a cause, a cause?
  - The role of theory
- RCT's, observational data and longitudinal analysis
  - The problems of cross-sectional, panel and cohort studies
- Change in quantity and change in states
- The 'difference in difference' approach
- The fixed effects approach
  - Demeaning
  - First differences

Time, Cause and Statistics

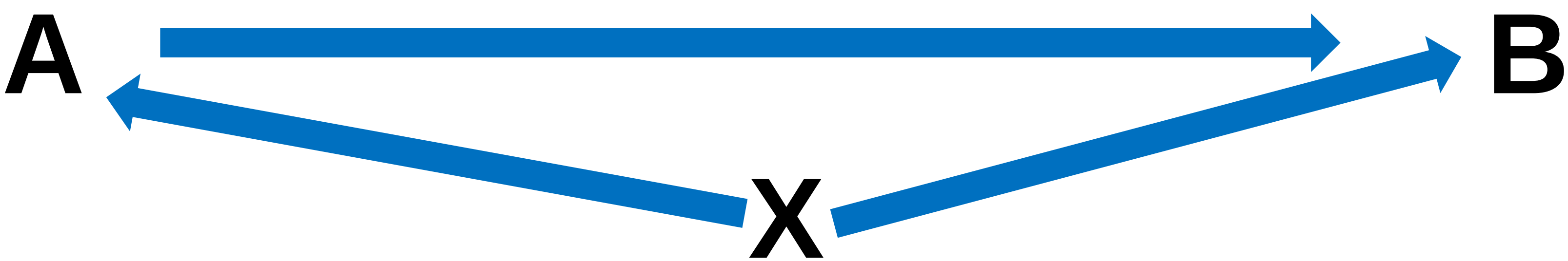
- Cause and effect are natural concepts
- Social scientists (part. Sociologists) can be queasy about ‘cause’
  - ‘Facts’ don’t speak for themselves
  - Invisible ‘social objects’ influence social behaviour
  - What are the causes of the causes?
    - Example, parenting in recession
- Empirical analysis needs to be guided by theory
- Have a model and hypotheses

Establishing Causality

- **The gold-standard of the randomised control trial:**
  - Isolation of ‘cause’ to one variable and manipulation
  - Random allocation to ‘intervention’ or ‘control’
  - Experimenter and subjects ‘blind’
- **Practical problems:**
  - Can we manipulate populations in natural settings
  - The issue of aggregates, e.g. communities
  - Lag periods, e.g. pensions, educational outcomes
- **Ethical problems:**
  - Can we expose individuals to ‘bads’?

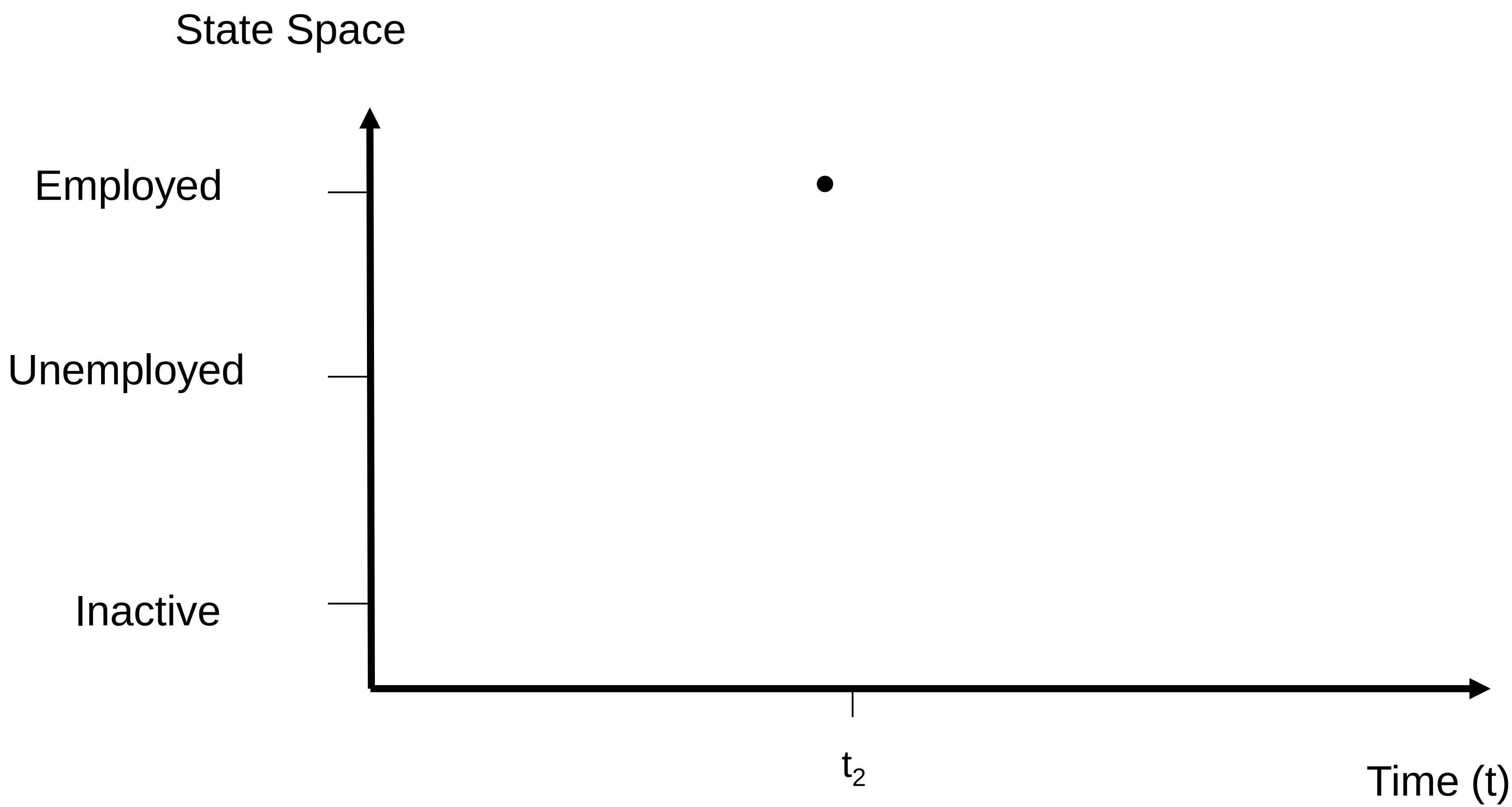
Longitudinal Studies

- The value of observational studies:
  - Naturalistic
  - Hopefully representative (variation)
- The problem of causality:



- Observing change over time – longitudinal studies
  - But how do we analyse these?

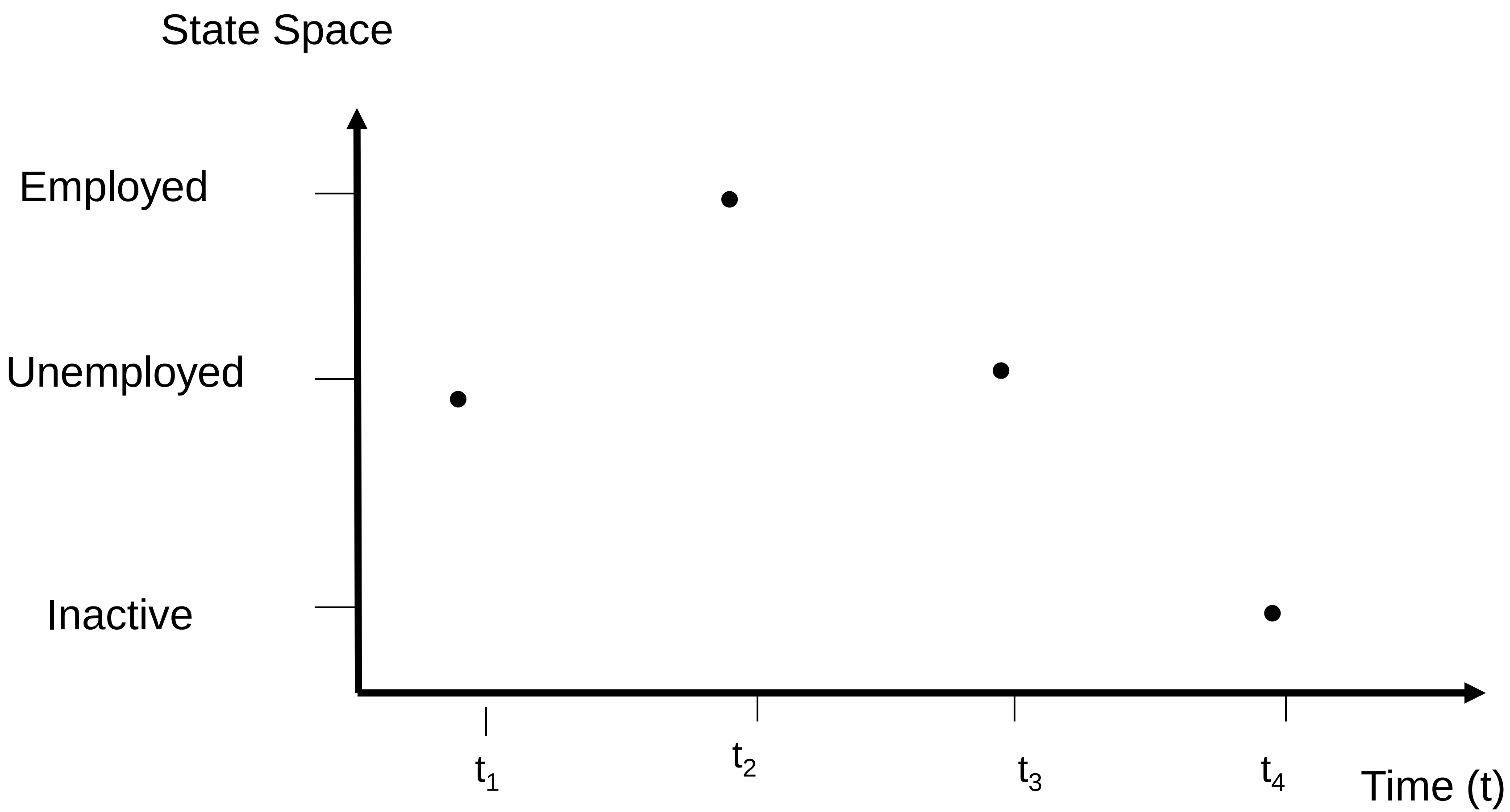
Cross-sectional Sample



Cross-Sectional Data

- CS analysis assume statistical equilibrium
- State probabilities are trend less and stable
- Coefficients express the net difference in the effects of predictors
- Cannot separate selection from causal effects
- Inferring causality is problematic – assume that predictors precede outcome and there are no feed backs

Data with Four Waves



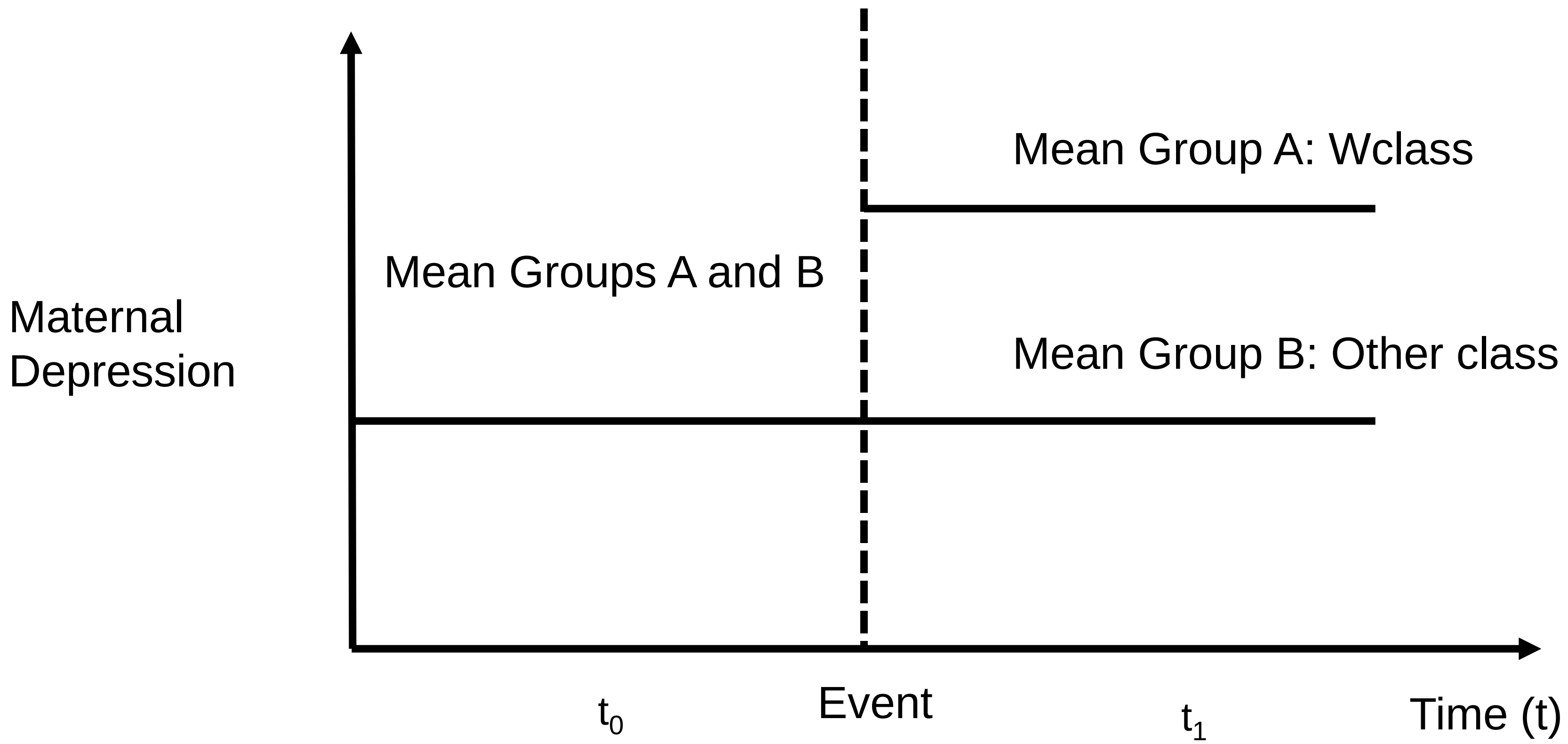
Change Over Two Time Points

- The addition of more observations permits analysis of quasi-causal relationships
- What predicts change from t_0 to t_1?
  - Change in whole population may just be coincidence
  - Change in a sub-population could be causal
- Leverage differential difference over time between groups: difference in difference
- Example: maternal depression, class and the impact of recession in 2010 compared to 2008

1. MD = α + Wclass + 2010 + error
2. MD = α + Wclass + 2010 + Wclass*2010 + error

Change Over Two Time Points

When analysing status change between two time points the classic approach is the 'difference in difference model':



Assumptions

- Outcome for both groups would be same in absence of ‘treatment’
- One group are not positively selected for the ‘treatment’
- If something else changed between t_0 and t_1, we can observe it and adjust for it
- There are no other, unobserved changes that will bias the estimate

Fixed Effects Models

- In the DinD estimate we assumed a great deal about changes that occurred between t_0 to t_1

$$MD = \alpha + Wclass + 2010 + Wclass{*}\underbrace{2010 + error}_{U_i}$$

- Multiple observations of the same individual allow us to control for individual differences
- Two approaches to ‘fixed-effects’:
    - Demeaning (sometimes called “within estimator)
    - First differencing

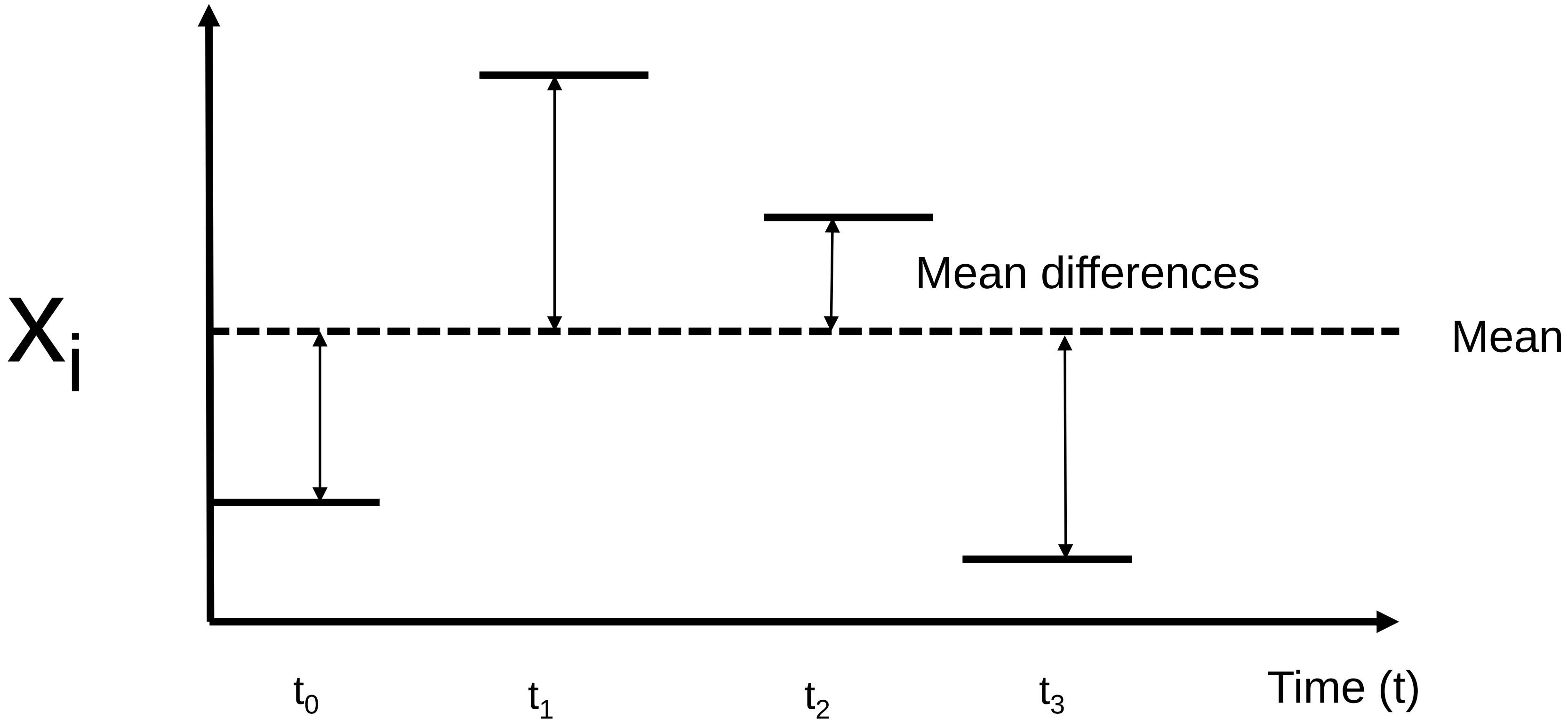
Demeaning

- With demeaning you (or computer) calculate individual averages of the dependent and explanatory variables
- You then subtract these averages from the regression equation:

$$\hat{y} = (yij - \bar{y}j) = \alpha + \beta(\chi ij - \bar{\chi}j) + \varepsilon ij$$

- So we are now predicting deviations from the individuals own mean rather than differences between individuals

Demeaning



First Differencing

- Alternative way of estimating the fixed effect is first differencing
- 'First differencing' subtracts the value of t_1 from the value of t_0 to produce the difference between the values
- We are thus explaining change at the individual level between periods
- Change can be explained by time constant (e.g. sex) and time-varying variables (e.g. income)
- First differencing can introduce serial correlation of the error terms so demeaning is usually a better option

Conclusion

- Longitudinal data offer a powerful tool for testing explanatory hypotheses
- Differential change across groups between periods useful even if the operative process unobserved
- Where operative variables observed, more powerful models can be estimated
- Explanatory analysis possible with two waves but more waves mean better estimates
- ‘Demeaning’ over 3 or more waves preferred