

The Research Value of Longitudinal Data

Professor Vernon Gayle

AQMen

University of Edinburgh

March 2019

© Vernon Gayle



Longitudinal Social Surveys

- Cross-sectional data
 - Respondents surveyed at only one time point
- Longitudinal data
 - Repeated contacts (with the same individuals)
 - Respondents surveyed at multiple time points

Research Using Longitudinal Social Survey Datasets

- For many social research projects cross-sectional data will be sufficient
- Most social research projects can be improved by the analysis of longitudinal data
- Some research questions require longitudinal data

A vignette...

The story of Jason Jones (aged 10) and his mum



Questions that Require Longitudinal Data

- Flows into and out of poverty
- The effects of family migration on the woman's subsequent employment activities
- Numerous policy intervention examples
- Numerous examples relating to 'individual' development

Methodological Benefits of Longitudinal Social Science Data

- Micro-level social processes
- Temporal ordering of events
- Improving control for residual heterogeneity
- Improving control for state dependence

Micro-Level Social Processes

- Cross-sectional data = a snap shot
 - Good for studying the immediate
 - Several datasets can study macro / or gross changes
- Repeated contacts data allow the study of
 - The passage of time
 - Individual (or household) change/stability
 - Processes that occur at the micro-level of the individual (or family)
 - Surprises (or shocks)

Temporal Ordering of Events (Direction of Influence)

- Time moves in one direction so...
 - An event in 1990 comes before an event in 1995
 - Experiences at primary school could affect university entry
 - Teenage smoking could influence health in old age
- But not *vice versa*

One sociology professor has argued with me suggesting that time does not move in only one direction

Temporal Ordering of Events (Direction of Influence)

- There is unequivocal evidence from cross-sectional data that, overall, the unemployed have poorer health
- This is consistent with both
 - A. Unemployment causing ill health
 - B. Ill health causing unemployment
- These two substantive stories are quite different

Month	Level of Health (20 = Good Health)	Employ Status
1	17	Employed
2	17	Employed
3	17	Employed
4	17	Unemployed
5	17	Unemployed
6	10	Unemployed
7	16	Unemployed
8	5	Unemployed
9	4	Unemployed
10	3	Unemployed
11	2	Unemployed
12	1	Unemployed

Person A



Became unemployed this has affected
his level of health

Month	Level of Health (20 = Good Health)	Employ Status
1	17	Employed
2	1	Employed
3	1	Employed
4	1	Unemployed
5	1	Unemployed
6	1	Unemployed
7	1	Unemployed
8	1	Unemployed
9	1	Unemployed
10	1	Unemployed
11	1	Unemployed
12	1	Unemployed

Person B



Poor health led to unemployment
(because of poor job performance)

In a cross-sectional study at month 12

- Person A would have been unemployed for 9 months and have a health score of 1
- Person B would have been unemployed for 9 months and have a health score of 1

This is an obvious example of how panel (i.e. repeated contacts) data can make an essential contribution to untangling social relationships

Improving Control for Omitted Explanatory Variables

- Residual Heterogeneity
 - Omitted explanatory variables
 - Unobserved heterogeneity
- The possibility of substantial variation between similar individuals due to unmeasured, and possibly immeasurable, variables is known as '*residual heterogeneity*'

Improving Control for Omitted Explanatory Variables

Because data collection instruments often fail to capture the detailed nature of social life there is, almost inevitably, considerable heterogeneity in response variables even amongst respondents that share the same characteristics across all of the explanatory variables

Improving Control for Omitted Explanatory Variables

As long as we make the assumption that (at least some of) these effects are enduring there are techniques for accounting for omitted explanatory variables if we have data at more than one time point

Improving Control for Omitted Explanatory Variables

- There are no routine methods of accounting for omitted explanatory variables in cross-sectional analysis
- It is sometimes claimed that the main advantage of longitudinal data is that it facilitates improved control for the plethora of variables that are omitted from any analysis
- Panel data won't completely sweep this problem away, but suitable models can improve control for, and estimate the effects of, residual heterogeneity

Improving Control for the Effects of Previous States (state dependence)

A frequently noted empirical regularity in the analysis of unemployment data is that those who were unemployed in the past or have worked in the past are more likely to be unemployed (or working) in the future

(Nobel Prize winner J.J. Heckman)

Improving Control for the Effects of Previous States (state dependence)

- Much of human behaviour is influenced by previous behaviour and outcomes (positive feedback)
- McGinnis (1968) '*axiom of cumulative inertia*'

Improving Control for the Effects of Previous States (state dependence)

- Working in May = more likely to be working in June
- Married this year = more likely to be married next year
- Own your own house this quarter
- Travel to work by car this week

Improving Control for the Effects of Previous States (state dependence)

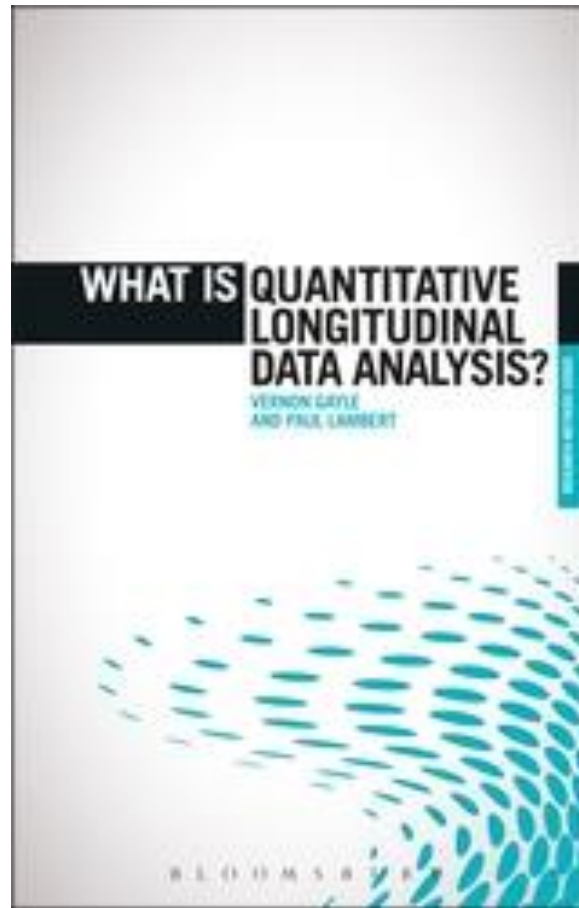
With panel data we may be able to include
past behaviour in the modelling process

Summary Message

There are methodological benefits...
but panel data are not a panacea!



Tweet - Longitudinal data enhance
our ability to investigate complicated
processes in the social world



Gayle, V. and Lambert, P. (2018) What is Quantitative Longitudinal Data Analysis? Bloomsbury Publishing.

Available at Amazon <https://tinyurl.com/y8cbf7c2>
Or directly from Bloomsbury <https://tinyurl.com/y9x5fltk>