



# AQMeN

Applied Quantitative Methods Network

## *Longitudinal Analysis*



UNIVERSITY OF STIRLING



AQMeN Introduction to Advanced Quantitative Techniques  
Thursday 13<sup>th</sup> May 2010

# Longitudinal Analysis

Professor Vernon Gayle  
University of Stirling

[vernon.gayle@stir.ac.uk](mailto:vernon.gayle@stir.ac.uk)

[www.longitudinal.stir.ac.uk](http://www.longitudinal.stir.ac.uk)

# Structure of this Presentation

- Introduction
- Longitudinal Data (the simple concept)
- Longitudinal Social Science Datasets
- Temporal Data Collection
- Data Collection Modes
- Longitudinal Data Structures
- Benefits of Longitudinal Data
- Longitudinal Models (15 years in 2 slides)
- Conclusions
- Where Next?

# Introduction

The focus of this talk is the analysis of longitudinal social survey data

- Cross-sectional data
  - Respondents surveyed at only one time point
- Longitudinal data
  - Repeated contacts
  - Respondents surveyed at multiple time points

# Introduction

- 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

# Introduction

- Some research questions require longitudinal data
  - Flows into and out of childhood poverty
  - The effects of family migration on the woman's subsequent employment activities
  - Numerous policy intervention examples
  - Numerous examples relating to 'individual' development

# Introduction

Longitudinal research is the lifeblood of the study of individual development. It has been pointed out many times that the most important questions concerning individual development can be answered only by applying a longitudinal design whereby the same individuals are followed through time.

(Bergman & Magnusson 1990)

# Longitudinal Data

Much of the time we are only interested in how some social phenomena affects a later outcome



We are using longitudinal data but standard cross-sectional techniques are still suitable

Analytically this is fairly trivial

# Longitudinal Data

Repeated outcome measures (one per contact)

<i>ID</i>	<i>Year</i>	<i>Age</i>	<i>Employment</i>
1	1991	16	Student
1	1992	17	Student
1	1993	18	Student
1	1994	19	Unemployed
1	1995	20	Employed (ft)
1	1996	21	Employed (ft)
1	1997	22	Employed (ft)
1	1998	23	Maternity Leave
1	1999	24	Family Care
1	2000	25	Employed (pt)

Analysis of repeated measures  
(i.e. multiple outcomes)  
is more complex

**Specialist techniques  
are required!**

# Longitudinal Social Science Datasets

- Micro social science datasets
  - Large  $n$  (e.g. BHPS 10k adults 5k households)
  - Small  $t$  (contacts once per year)
  - Large  $x$  (hundreds of variables)
- Contrast with time-series datasets
  - Unemployment rates; inflation; share prices etc.
  - Macro level (often countries)
  - Small  $n$  (one country; 10 EU states)
  - Frequent contacts (months by month for many years)
  - Few variables, sometime just one (e.g. inflation)

# Longitudinal Social Science Datasets

Two main forms of micro social longitudinal datasets

## 1 Panel Dataset

- Repeated contacts data collection
- Sociologist Paul Lazarsfeld opinion research in 1930s
- Common example is the Household Panel Study

# Longitudinal Social Science Datasets

- Cohort Study
  - Repeated contacts data collection
  - Principally concerned with charting the development of a particular ‘group’ from a certain point in time
  - (simply a specific form of panel design in my view)
  
  - A birth cohort of babies born in a particular year
  - A youth cohort, a group of pupils who completed compulsory education in the same year
  - A group of newly qualified doctor

# Longitudinal Social Science Datasets

- Panel Dataset Examples (Household Panel Studies)
  - US Panel Study of Income Dynamics (PSID)
    - began in 1968 <http://psidonline.isr.umich.edu/>
  - Germany Socio-Economic Panel (SOEP)
    - began in 1984 <http://www.diw.de/en/soep>
  - British Household Panel Survey BHPS (1991 onwards)
    - 5k households, 10k adults, <http://www.iser.essex.ac.uk/survey/bhps>

# Longitudinal Social Science Datasets

## New British Panel Survey

- Understanding Society (US)
  - Also known as the UK Household Longitudinal Study (UKHLS)
- Began in January 2009
- Incorporates and extends the BHPS
- 40k UK households (4k Scottish Households)
- 4k households in a special ethnic minorities sample
- Innovations include
  - linking to administrative data; spatial data; biometric data; qualitative data; child data (from age 10)
- <http://www.understandingsociety.org.uk/>

# Longitudinal Social Science Datasets

- Cohort Dataset Examples
  - Birth Cohort Studies
    - 1946, 1958, 1970 & Millennium Cohort Study
    - <http://www.cls.ioe.ac.uk/>
    - <http://www.nshd.mrc.ac.uk/>
  - Youth Cohort Study of England and Wales (YCS)
    - <http://www.data-archive.ac.uk/findingData/ycsTitles.asp>
  - BMA Cohort Study of Newly Qualified Doctors
    - [http://www.bma.org.uk/healthcare\\_policy/cohort\\_studies/index.jsp](http://www.bma.org.uk/healthcare_policy/cohort_studies/index.jsp)

# Longitudinal Social Science Datasets

- Scottish Datasets Examples

- Growing Up in Scotland (GUS)

- A birth cohort study of 8k children
    - <http://www.crfr.ac.uk/gus/>

- Scottish Longitudinal Study

- A panel study of 274k people based on Census records
    - <http://www.lscs.ac.uk/sls/>

# Temporal Data Collection

Prospective



Retrospective



Most studies collect a mixture of both

# Data Collection Modes

- Multiple modes are common longitudinal studies
- Standard interviews
- Traditional questionnaires
- Computer aided interviews
- Telephone interviews
- Increasingly innovative electronic data collection

# Longitudinal Data Structures


A simple example of a panel (repeated contacts) dataset

<i>ID</i>	<i>Year</i>	<i>Age</i>	<i>Gender</i>	<i>Employment</i>	<i>Marital Status</i>
1	1991	16	Female	Student	Single
1	1992	17	Female	Student	Single
1	1993	18	Female	Student	Single
1	1994	19	Female	Unemployed	Single
1	1995	20	Female	Employed (ft)	Cohabiting
1	1996	21	Female	Employed (ft)	Cohabiting
1	1997	22	Female	Employed (ft)	Cohabiting
1	1998	23	Female	Maternity Leave	Married
1	1999	24	Female	Family Care	Married
1	2001	25	Female	Employed (pt)	Separated

# Longitudinal Data Structures

“Long” format dataset


<i>ID</i>	<i>Year</i>	<i>Age</i>	<i>Gender</i>	<i>Employment</i>	<i>Marital Status</i>
11991	16				
1	1992	17			
1	1993	18			
1	1994	19			
1	1995	20			
1	1996	21			
1	1997	22			
11998	23				
1	1999	24			
1	2001	25			



# Longitudinal Data Structures

Repeated outcome measures (one per contact)


<i>ID</i>	<i>Year</i>	<i>Age</i>	<i>Gender</i>	<i>Employment</i>	<i>Marital Status</i>
1	1991	16	Female	Student	Single
1	1992	17	Female	Student	Single
1	1993	18	Female	Student	Single
1	1994	19	Female	Unemployed	Single
1	1995	20	Female	Employed (ft)	Cohabiting
1	1996	21	Female	Employed (ft)	Cohabiting
1	1997	22	Female	Employed (ft)	Cohabiting
1	1998	23	Female	Maternity Leave	Married
1	1999	24	Female	Family Care	Married
1	2001	25	Female	Employed (pt)	Separated



# Longitudinal Data Structures

Time constant explanatory variables


<i>ID</i>	<i>Year</i>	<i>Age</i>	<i>Gender</i>	<i>Employment</i>	<i>Marital Status</i>
1	1991	16	Female		Single
1	1992	17	Female		Single
1	1993	18	Female		Single
1	1994	19	Female		Single
1	1995	20	Female		Cohabiting
1	1996	21	Female		Cohabiting
1	1997	22	Female		Cohabiting
1	1998	23	Female		Married
1	1999	24	Female		Married
1	2001	25	Female		Separated



# Longitudinal Data Structures

Time changing explanatory variables

<i>ID</i>	<i>Year</i>	<i>Age</i>	<i>Gender</i>	<i>Employment</i>	<i>Marital Status</i>
1	1991	16	Female	Student	Single
1	1992	17	Female	Student	Single
1	1993	18	Female	Student	Single
1	1994	19	Female	Unemployed	Single
1	1995	20	Female	Employed (ft)	Cohabiting
1	1996	21	Female	Employed (ft)	Cohabiting
1	1997	22	Female	Employed (ft)	Cohabiting
1	1998	23	Female	Maternity Leave	Married
1	1999	24	Female	Family Care	Married
1	2001	25	Female	Employed (pt)	Separated



# Longitudinal Data Structures

Time to an event

Time to first childbirth



ID=1

# Benefits of Longitudinal Data

- Some research questions require longitudinal data
  - Micro-level change over time
    - Flows into and out of childhood poverty
    - The effects of family migration
    - Policy intervention examples
    - ‘Individual’ development

# Benefits of Longitudinal Data

- Additional methodological benefits
  - Temporal ordering of events (direction of causality)
  - Improved control for omitted explanatory variables (residual heterogeneity)
  - Improved control for the effects of previous states (state dependence)
  - Exploring the effects of both ageing and cohort membership (age-period-cohort effects)

# Longitudinal Models

- Two main modelling approaches in social science research
  1. Event history analysis, time to an event
    - Also known as duration analysis; survival analysis; failure time analysis; duration economics; hazard modelling

Generally time is continuous and we model the probability of an event occurring given that it has not already occurred (hazard)

# Longitudinal Models

- Two main modelling approaches in social science research

## 2. Panel data analysis

- Regression models suitable for repeated observations
- Time generally conceptualised as being discrete
- Extension of standard regression models (glm)
- Closely related to multilevel modelling (glmm)
- More advanced versions (e.g. dynamic models)
  
- Alternative terminology
  - variance components models; hierarchical linear models; cross-sectional time series; random effects modelling

# Conclusion

- 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

# Conclusion

- Longitudinal are not a panacea but data facilitate
  - The study of micro-level change social change over time (and also social stability)
  - A better understanding of the temporal ordering of events (direction of causality)
  - Improved control for omitted explanatory variables (residual heterogeneity)
  - Improved control for the effects of previous states (state dependence)
  - Exploration of the effects of both ageing and cohort membership (age-period-cohort effects)

# Where Next?

- More advanced skills are required
  - Extension from standard modelling techniques
  - Software (I recommend Stata)
- [www.longitudinal.stir.ac.uk](http://www.longitudinal.stir.ac.uk)
- Annotated reading list  
[http://www.longitudinal.stir.ac.uk/refs/reading\\_lda\\_08.pdf](http://www.longitudinal.stir.ac.uk/refs/reading_lda_08.pdf)
- AQMeN training planned for early 2011