Quantitative Approaches to Longitudinal Research

Paul Lambert
Dept Applied Social Science, Stirling University
paul.lambert@stir.ac.uk
http://staff.stir.ac.uk/paul.lambert/

Vernon Gayle
as pl. and: Longitudinal Studies Centre, University of St Andrews
vernon.gayle@stirling.ac.uk

Session 2 of RCBN training workshop
Longitudinal Research in Education,
University of York, 23.1.04

Quantitative Approaches to Longitudinal Research

1. Introduction to and motivations for quantitative longitudinal research
2. Repeated cross-sections
3. Panel datasets
4. Cohort studies
5. Event history datasets
6. Time series analyses

Introduction to quantitative longitudinal research

i. What constitutes QnLR?
ii. Hypothetical motivations for QnLR
iii. (Perceived) Pragmatic Drawbacks to QnLR
iv. Some advice…

i) QnLR in the social sciences

• Data analysis ⇨ Survey resources
  – Micro-data (individuals, households, ..)
  – Macro-data (aggregate summary for year, country..)

Data analysis is used to give a parsimonious summary of patterns of relations between variables in the survey dataset

i) QnLR in the social sciences

• Qn longitudinal research involves
  – Focus on time / durations
  – Focus on change (or stability)
  – Use of time as control: ‘residual heterogeneity’

• Specific features to QnLR
  – Tends to use ‘large and complex’ secondary datasets
  – Particular techniques of data analysis
Large & complex social science data
⇒ Complexity in:
- Multiple points of measurement
- Multiple hierarchies of measurement
- Array of variables / operationalisations
- Relations between / subgroups of cases
- Sample collection and weighting

ii) Motivations for QnLR
- Substantive role of time dimension:
  - repeated information over time
  - durations
  - changes and stability
- Distinguish age, period and cohort effects
- Causality and residual heterogeneity
- Examining multivariate relationships
- Representative conclusions
- Secondary data analysis positives: other users; cheap access; range of topics available

iii) Drawbacks to QnLR
- Dataset expense: mostly 2ndry; limited access to some data (cf disclosure risk)
- Data analysis: software needs due to complexity of some methods
- Unavoidable data management: complex file & variable management requires training and skills of good practice

iv) QnLR: Prescriptions for the casual user
- Be aware of range of analytical options
  - Some research fields get stuck in paradigms
- Technical skills = data management
  - Put some effort into understanding software command files (eg STATA or SPSS syntax; ‘dipping into’ windows menus isn’t enough)
- Network
  - UK QnLR social sciences is quite a small field, currently quite supportive to each other

QnLR: Some research resources
- See reading list for text and internet resources
- Training courses in UK – see RCBN site
- A plug for 2 projects at Stirling:
  - Longitudinal Studies Seminar Series
    http://www.lss.stir.ac.uk
    eg workshop St Andrews 27/8 Feb
  - ‘Longitudinal studies in the social sciences’
    ESRC funded research methods programme, e-learning & workshops 2004/5, contact vernon.gayle@stirling.ac.uk

Quantitative Approaches to Longitudinal Research
1. Introduction to and motivations for quantitative longitudinal research
2. Repeated cross-sections
3. Panel datasets
4. Cohort studies
5. Event history datasets
6. Time series analyses
Repeated Cross-sections

- **By far the most widely used** longitudinal analysis in contemporary social sciences

Whole surveys, with same variables, repeated at different time points

and

Same information extracted from different surveys from different time points

Illustration: Repeated x-sect data

<table>
<thead>
<tr>
<th>Survey</th>
<th>Person</th>
<th>Person-level Vars</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1 38 1 1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2 34 2 2</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>2 6 - -</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>1 45 1 3</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>2 41 1 1</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>1 20 2 2</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>1 25 2 2</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>1 20 1 1</td>
</tr>
</tbody>
</table>

N s=3 N c=8

Some leading repeated cross-section surveys: UK

<table>
<thead>
<tr>
<th>OPCS Census</th>
<th>British Crime Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labour Force Survey</td>
<td>British Social Attitudes</td>
</tr>
<tr>
<td>New Earnings Survey</td>
<td>British Election Studies</td>
</tr>
<tr>
<td>Family Expenditure Survey</td>
<td>Policy Studies (Ethnicity)</td>
</tr>
<tr>
<td>General Household Survey</td>
<td>Social Mobility enquiries</td>
</tr>
</tbody>
</table>

Some leading repeated cross-section surveys: International

<table>
<thead>
<tr>
<th>European Social Survey</th>
<th>PISA (schoolkid’s aptitudes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPUMS census harmonisation</td>
<td>ISSP</td>
</tr>
<tr>
<td>LIS/LES (income and employment)</td>
<td>Eurobarometer</td>
</tr>
</tbody>
</table>

Repeated cross sections

- **Easy to communicate** & appealing: how things have changed between certain time points
- Distinguishes **age / period / cohort**
- Easier to analyse – **less data management**

   However...

   ☺ Don’t get other QnLR attractions (nature of changers; residual heterogeneity; causality; durations)

   ☺ Hidden complications: are sampling methods, variable operationalisations really comparable? (don’t overdo: concepts are more often robust than not)

Repeated X-sectional analysis

1. **Present stats distinctively by time pts**
   - Analytically sound
   - Tends to be descriptive, limited # vars

2. **Time points as an explanatory variable**
   - More complex, requires more assumptions of data comparability
   - Can allow a more detailed analysis / models
Example 2.1: UK Census

- Directly access aggregate statistics from census reports, books or web, eg:

<table>
<thead>
<tr>
<th>Wales: Proportion able to speak Welsh</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
</tr>
</tbody>
</table>

- Census not that widely used: larger scale surveys often more data and more reliable

Example 2.2i: LFS yearly stats

Percent of UK workers with a higher degree, by employment category and gender (m/f)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher degree</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>2.383</td>
<td>.000</td>
<td>10.842</td>
</tr>
<tr>
<td>Time point 1991</td>
<td>-0.955</td>
<td>.000</td>
<td>0.385</td>
</tr>
<tr>
<td>Time point 2001</td>
<td>-4.232</td>
<td>.000</td>
<td>0.015</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.339</td>
<td>.000</td>
<td>0.717</td>
</tr>
<tr>
<td>(Time in years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nagelkere R2</td>
<td>.11</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Eg2.2: UK Labour Force Survey

see syntax examples 2.2

LFS: free download from UK data archive
http://www.data-archive.ac.uk/

Same questions asked yearly / quarterly

i. Analysis on separate time groups

ii. Analysis by time: pool records and use time and interactions with time as explanatory variables

Example 2.2ii: LFS and time


<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>Sig.</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher degree</td>
<td>2.383</td>
<td>.000</td>
<td>10.842</td>
</tr>
<tr>
<td>Female</td>
<td>-0.955</td>
<td>.000</td>
<td>0.385</td>
</tr>
<tr>
<td>Age in years /10</td>
<td>0.777</td>
<td>.000</td>
<td>2.174</td>
</tr>
<tr>
<td>Age squared /1000</td>
<td>-1.897</td>
<td>.000</td>
<td>0.184</td>
</tr>
<tr>
<td>Time point 1991</td>
<td>0.094</td>
<td>.000</td>
<td>1.098</td>
</tr>
<tr>
<td>Time point 2001</td>
<td>-1.496</td>
<td>.000</td>
<td>0.231</td>
</tr>
<tr>
<td>(Time in years)* (Higher Degree)</td>
<td>-0.339</td>
<td>.000</td>
<td>0.717</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.336</td>
<td>.000</td>
<td>0.717</td>
</tr>
</tbody>
</table>

Quantitative Approaches to Longitudinal Research

1. Introduction to and motivations for quantitative longitudinal research
2. Repeated cross-sections
3. Panel datasets
4. Cohort studies
5. Event history datasets
6. Time series analyses

Panel Datasets

Information collected on the same cases at more than one point in time

- ‘classic’ longitudinal design
- incorporates ‘follow-up’, ‘repeated measures’, and ‘cohort’
Panel data in the social sciences

- **Large scale studies** are ambitious and expensive, normally by major organisations, with efforts made to promote use
- But **small scale panels** are surprisingly common...
- ‘Balanced’ and ‘Unbalanced’ designs

### Illustration: Unbalanced panel

<table>
<thead>
<tr>
<th>Wave</th>
<th>Person</th>
<th>← Person-level Vars →</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-</td>
</tr>
</tbody>
</table>

N_w=3, N_p=3, *also ‘sweep’, ‘contact’...

Panel data advantages

- **Study ‘changers’** – how many of them, what are they like, what caused change
- Control for **individuals’ unknown characteristics** (‘residual heterogeneity’)
- Develop a full and **reliable life history** – eg family formation, employment patterns

Panel data drawbacks

- **Data analysis**: can be complex, some methods are advanced / still developing
- **Data management**: tends to complexity, need training to get on top of
- **Dataset access**: varies by topics (disclosure risks)
- **Attrition** (dropout of cases from recontacts)
- **Long Duration**: eg politics of funding

Some leading panel surveys: UK

- British Household Panel Study (BHPS)
- ONS Longitudinal Study (Census 1971-)
- British Election Panel Studies
- Labour Force Survey rotating panel
- School attainment studies (various)
- Health and medical progress studies (various)

Some leading panel studies: International

- European Community Household Panel Study
- CHER, PACO, CNEF (individual projects harmonising panels)
- Panel Study of Income Dynamics (US)
Analytical approaches

i) Study of Transitions / changers
   - simple methods in any package, eg cross-tab if changed or not by background influence
   - but complex data management

ii) Study of durations / life histories
   - See section 5 ‘event histories’

Panel data model types

- Fixed effects - suits smaller samples
- Random effects - ‘variance components’, ‘multilevel model’, suits larger samples
- Growth curves - as VC + time parameters
- Dynamic Lag-effects models - best in theory, but most complex methodologically
  Analytically complex and often need advanced or specialist software, eg STATA, S-PLUS, SABRE/GLIM, LIMDEP, MLWIN, ...

Example 3.1: Panel transitions

- BHPS: annual survey of ~ 10,000 adult householders, plus children interviews
- Attrition very low (though initial non-contacts, & item non-response, are higher)
- Complex recent sample additions
- See many promotional websites...

Aside: Major BHPS issue is successful data management

- Good practice:
  - SPSS or STATA command files to match records between multiple data files
  - Liaise with BHPS user support at Essex
  - Attend training workshops / see other users’ files, eg http://www.cf.ac.uk/socsi/main/lambertp/stirbhp

Example 3.1: Panel transitions

Young people’s household circumstance changes by subjective well-being between 1994 and 1995.
BHPS youth panel, 11-14yrs in 1994, row percent.

<table>
<thead>
<tr>
<th></th>
<th>Stays happy</th>
<th>Cheers up</th>
<th>Becomes miserable</th>
<th>Stays miserable</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>HH Stable</td>
<td>54%</td>
<td>19%</td>
<td>10%</td>
<td>18%</td>
<td>499</td>
</tr>
<tr>
<td>HH Changes</td>
<td>42%</td>
<td>22%</td>
<td>14%</td>
<td>22%</td>
<td>81</td>
</tr>
</tbody>
</table>
Example 3.2: Panel model

BHPS 1994-8: Output from Variance Components Panel model for determinants of GHQ scale score (higher = more miserable), by individual factors for multiple time points per person.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>12.69</td>
<td>.168</td>
<td>.000</td>
<td>12.4</td>
<td>13.0</td>
</tr>
<tr>
<td>Female</td>
<td>-1.36</td>
<td>.075</td>
<td>.000</td>
<td>-1.5</td>
<td>-1.2</td>
</tr>
<tr>
<td>In work</td>
<td>-1.23</td>
<td>.082</td>
<td>.000</td>
<td>-1.4</td>
<td>-1.1</td>
</tr>
<tr>
<td>Unemployed</td>
<td>.50</td>
<td>.131</td>
<td>.000</td>
<td>2</td>
<td>.8</td>
</tr>
<tr>
<td>FT studying</td>
<td>-1.70</td>
<td>.141</td>
<td>.000</td>
<td>-2.0</td>
<td>-1.4</td>
</tr>
<tr>
<td>Age in years</td>
<td>.00</td>
<td>.002</td>
<td>.055</td>
<td>.0</td>
<td>.0</td>
</tr>
<tr>
<td>Holds degree or diploma</td>
<td>.07</td>
<td>.076</td>
<td>.356</td>
<td>-2</td>
<td>.1</td>
</tr>
<tr>
<td>Time point</td>
<td>.03</td>
<td>.014</td>
<td>.020</td>
<td>.0</td>
<td>.1</td>
</tr>
</tbody>
</table>

95% Confidence Interval

Variance components: Person level = 46%, individual level = 54%.

Quantitative Approaches to Longitudinal Research

1. Introduction to and motivations for quantitative longitudinal research
2. Repeated cross-sections
3. Panel datasets
4. Cohort studies
5. Event history datasets
6. Time series analyses

Cohort Datasets

Information on a group of cases which share a common circumstance, collected repeatedly as they progress through a life course

- Simple extension of panel dataset
- One of most widely used and intuitive types of repeated contact data (eg ‘7-up’ series)

Cohort data in the social sciences

- Circumstances parallel other panel types:
  - Large scale studies ambitious & expensive
  - Small scale cohorts still quite common…
  - Attrition problems often more severe
  - Considerable study duration problems – have to wait for generations to age

Cohort data advantages

- Study of ‘changers’ a main focus, looking at how group of cases develop after a certain point in time
- Build up a full and reliable life history – often covers a very long span
- May inform on a variety of issues as cohort progresses through lifecourse

Cohort data drawbacks

- Data analysis & management complexity (less severe than other panels, as generally fewer contact points)
- Attrition problems more severe than panel
- Longer Duration
- Very specific findings – eg only for people of a specific age group
Some leading UK cohort surveys

**Birth Cohort Studies**
- 1946 National Survey of Health and Development
- 1958 National Child Development Study
- 1970 Birth cohort study
- 2000 Millenium Cohort Study

**Youth Cohort Studies (1985 onwards)**
- Health and medical progress studies (various)
- Criminology studies of recidivism (various)

**Example cohort dataset:**
1958 National Child Development Study
- All women who gave birth in G.B. during the week of 3-9th March 1958.
- Approx 17,000 original subjects
- Several contacts for perinatal mortality studies
- Subsequent ‘sweeps’ of children at ages 7, 11, 16, 20, 23, 33, 41
- 12% attrition by 1991 age 33 (this is quite low...)

Cohort data analytical approaches
- parallel those of other panel data:
  i. Study of transitions / changers
  ii. Study of durations / life histories
  iii. Panel data models

*But focus more on life-course development than shorter term transitions*

Cohort data analysis example:
“Econometric Analysis of the Demand for Higher Education” (Gayle, Berridge and Davies 2003)
- **Youth Cohort Study** annual recontacts with teenagers as progress through, then leave, school
- **Statistical models:**
  - chances of staying to A-levels,
  - then chances of University start, given A-level route

Joint modelling process shows:

<table>
<thead>
<tr>
<th>Predictors of A-level route at 16:</th>
<th>Predictors of Univ. route at 18, given A-level at 16:</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCSE score</td>
<td>GCSE score, A-level score; Indian, Chinese, O. Asian</td>
</tr>
<tr>
<td>Indian</td>
<td></td>
</tr>
<tr>
<td>Parental class</td>
<td></td>
</tr>
<tr>
<td>Parents’ education</td>
<td></td>
</tr>
<tr>
<td>Home tenure</td>
<td></td>
</tr>
<tr>
<td>School type</td>
<td></td>
</tr>
<tr>
<td>Region in UK</td>
<td></td>
</tr>
</tbody>
</table>

Quantitative Approaches to Longitudinal Research

1. Introduction to and motivations for quantitative longitudinal research
2. Repeated cross-sections
3. Panel datasets
4. Cohort studies
5. Event history datasets
6. Time series analyses
Event history data analysis

Focus shifts to length of time in a ‘state’ - analyses determinants of time in state

- Alternative data sources:
  - Panel / cohort (more reliable)
  - Retrospective (cheaper, but recall errors)
- Aka: ‘Survival data analysis’; ‘Failure time analysis’; ‘hazards’; ‘risks’; ..

Social Science event histories:

- Time to labour market transitions
- Time to family formation
- Time to recidivism

Comment: Data analysis techniques relatively new, and aren’t suited to complex variates

Many event history applications have used quite simplistic variable operationalisations

Event histories differ:

- In form of dataset (cases are spells in time, not individuals)
  - Some complex data management issues
- In types of analytical method
  - Many techniques are new or rare, and specialist software may be needed

Key to event histories is ‘state space’

Episodes within state space: Lifetime work histories for 3 adults born 1935

| State space | Person 1 | | | | | | |
|------------|----------|-----------------|-----------------|-----------------|-----------------|-----------------|
| FT work    | FT work  | FT work         | FT work         | FT work         | FT work         | FT work         |
| FT work    | FT work  | FT work         | FT work         | FT work         | FT work         | FT work         |
| FT work    | FT work  | FT work         | FT work         | FT work         | FT work         | FT work         |

| State space | Person 2 | | | | | | |
|------------|----------|-----------------|-----------------|-----------------|-----------------|-----------------|
| FT work    | FT work  | FT work         | FT work         | FT work         | FT work         | FT work         |
| FT work    | FT work  | FT work         | FT work         | FT work         | FT work         | FT work         |
| FT work    | FT work  | FT work         | FT work         | FT work         | FT work         | FT work         |

| State space | Person 3 | | | | | | |
|------------|----------|-----------------|-----------------|-----------------|-----------------|-----------------|
| FT work    | FT work  | FT work         | FT work         | FT work         | FT work         | FT work         |
| FT work    | FT work  | FT work         | FT work         | FT work         | FT work         | FT work         |
| FT work    | FT work  | FT work         | FT work         | FT work         | FT work         | FT work         |

Illustration of a continuous time retrospective dataset

<table>
<thead>
<tr>
<th>Case</th>
<th>Person</th>
<th>Start time</th>
<th>End time</th>
<th>Duration</th>
<th>Origin State</th>
<th>Destination State</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>158</td>
<td>157</td>
<td>FT work</td>
<td>NW</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>158</td>
<td>170</td>
<td>12</td>
<td>PT work</td>
<td>NW</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>22</td>
<td>196</td>
<td>3</td>
<td>PT work</td>
<td>NW</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>22</td>
<td>196</td>
<td>5</td>
<td>PT work</td>
<td>NW</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>149</td>
<td>170</td>
<td>21</td>
<td>PT work</td>
<td>NW</td>
</tr>
</tbody>
</table>

Illustration of a discrete time retrospective dataset

<table>
<thead>
<tr>
<th>Case</th>
<th>Person</th>
<th>Discrete Time</th>
<th>Approx real time</th>
<th>State End of state</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>FT 0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>20</td>
<td>FT 0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>35</td>
<td>FT 0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>4</td>
<td>50</td>
<td>FT 0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>5</td>
<td>65</td>
<td>FT 0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>6</td>
<td>80</td>
<td>FT 0</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>7</td>
<td>95</td>
<td>FT 0</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>8</td>
<td>110</td>
<td>FT 0</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>9</td>
<td>125</td>
<td>FT 0</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>10</td>
<td>140</td>
<td>FT 1</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>11</td>
<td>155</td>
<td>NW 0</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>12</td>
<td>170</td>
<td>NW 3</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>NW 0</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>2</td>
<td>20</td>
<td>NW 1</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>3</td>
<td>35</td>
<td>FT 0</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>4</td>
<td>50</td>
<td>FT 1</td>
</tr>
</tbody>
</table>

[Page numbers are provided for reference.]
Event history data permutations

- **Single state single episode**
  - Eg Duration in first post-school job till end

- **Single episode competing risks**
  - Eg Duration in job until promotion / retire / unemp.

- **Multi-state multi-episode**
  - Eg adult working life histories

- **Time varying covariates**
  - Eg changes in family circumstances as influence on employment durations

Some UK event history datasets

- British Household Panel Study (see separate 'combined life history' files)
- National Birth Cohort Studies
- Family and Working Lives Survey
- Social Change and Economic Life Initiative
- Youth Cohort Studies

Event history analysis software

- **SPSS** – very limited analysis options
- **STATA** – wide range of pre-prepared methods
- **SAS** – as STATA
- **S-Plus/R** – vast capacity but non-introductory
- **GLIM / SABRE** – some unique options
- **TDA** – simple but powerful freeware
- **MLwiN; IEM; {others}** – small packages targeted at specific analysis situations

Types of Event History Analysis

i. **Descriptive**: compare times to event by different groups (eg survival plots)

ii. **Modelling**: variations of Cox’s Regression models, which allow for particular conditions of event history data structures

  - Type of data permutations influences analysis – only simple data is easily used!

Eg 5.1: Mean durations by states

BHPS first job durations by EGP class

![Graph showing mean durations by states for BHPS first job durations by EGP class.]

Eg 5.1: Kaplan-Meir survival

BHPS males 1st job KM

![Graph showing Kaplan-Meir survival analysis for BHPS males 1st job.]

[Diagram showing Kaplan-Meir survival analysis for BHPS males 1st job.]

- agricultural wks
- service class hi
- routine non-med
- service class lo
- personal service
- small props w/o
- farmers
Eg 5.2: Cox’s regression

Cox regression estimates: risks of quicker exit from first employment state of BHPS adults

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>SE</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>.194</td>
<td>.081</td>
<td>.017</td>
</tr>
<tr>
<td>Self-employed</td>
<td>-.617</td>
<td>.179</td>
<td>.001</td>
</tr>
<tr>
<td>Age in 1990</td>
<td>-.062</td>
<td>.003</td>
<td>.000</td>
</tr>
<tr>
<td>Age in 1990 squared</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>Hope-Goldthorpe scale</td>
<td>-.013</td>
<td>.001</td>
<td>.000</td>
</tr>
<tr>
<td>Female*Self-employed</td>
<td>.214</td>
<td>.109</td>
<td>.049</td>
</tr>
<tr>
<td>Female*HG scale</td>
<td>-.003</td>
<td>.002</td>
<td>.061</td>
</tr>
<tr>
<td>Self-employed*HG scale</td>
<td>.000</td>
<td>.004</td>
<td>.897</td>
</tr>
<tr>
<td>Female*Age in 1990</td>
<td>.006</td>
<td>.001</td>
<td>.000</td>
</tr>
</tbody>
</table>

Quantitative Approaches to Longitudinal Research

| 1. Introduction to and motivations for quantitative longitudinal research | 2. Repeated cross-sections |
| 3. Panel datasets | 4. Cohort studies |
| 5. Event history datasets | 6. Time series analyses |

Time series data

Statistical summary of one particular concept, collected at repeated time points from one or more subjects

Examples:
- Unemployment rates by year in UK
- University entrance rates by year by country

Time Series Analysis

i) Descriptive analyses
- charts / text commentaries on values by time periods and different groups
- Widely used in social science research
- But exactly equivalent to repeated cross-sectional descriptives.

Time Series Analysis

ii) Time Series statistical models
- Advanced methods of modelling data analysis are possible, require specialist stats packages
- Main feature: many measurement points and few cases
- Major area in business / economics, but largely unused in other social sciences

Some UK Time Series sources

- Time series databases (aggregate statistics)
  - ONS Time series data
  - ESDS International macrodata

- Repeated cross-sectional surveys
  - Census
  - Labour Force Survey
  - Many others..
…Phew!

Summary: Quantitative approaches to longitudinal research

1) Pro’s and cons to QnL research:

i. Appealing analytical possibilities: eg analysis of change, controls for residual heterogeneity

ii. Pragmatic constraints: data access, management, & analytical methods; often applications over-simplify variables

iii. Uneven penetration of research applications between research fields at present

2) Importance of knowing the research field:

i. Don’t have to be an expert: On longitudinal research influential to social sciences – should be able to understand / critique even if not a practitioner

ii. Understand data collection process and retain sociological background

iii. Make better research proposals!

3) Undertaking QnL research:

i. Needs a bit of effort: learn software, data management practice – workshops and training facilities available

ii. Remain sociologically driven: ‘methodolatry’ widespread in QnL, applications ‘forced’ into desired techniques; often simpler techniques make for the more popular & influential reports

iii. Learn by doing (…try the syntax examples…)

See reading list / internet resources for further information and

http://staff.stir.ac.uk/paul.lambert/rcbn_york04/