Individuals in Household Panels: 
The importance of person group clustering

Paul Lambert & Vernon Gayle

Dept. Applied Social Science, University of Stirling, and ISER, University of Essex

Paper presented to the session ‘Analysis of Panel Data Based on Complex Longitudinal Surveys’, ISA RC33 7th International Conference on Social Science Methodology, Naples, 1-5 September 2008
The British Household Panel Study 1991->

- Panel study of individuals from 5.5k households contacted in 1991, re-contacted annually
- Major UK research investment
- Incorporation into ‘UK Household Longitudinal Study’ (UKHLS) 2008 ->

For lots more introductions, see: http://www.longitudinal.stir.ac.uk/
BHPS Sampling design

• W1 (1991): Stratified random sample of 5,500 households
  ➢ 14,000 ‘OSM’ household members
  ➢ Later waves: trace all OSM’s; their descendants; and their household sharers (TSM’s\PSM’s); (and ‘boost’ samples)

Longitudinal trace of individuals and their surrounding household, but not of ‘longitudinal households’
<table>
<thead>
<tr>
<th>Wave:</th>
<th>OSM (inc PSMs)</th>
<th>TSM (essex)</th>
<th>ECHP boost</th>
<th>Scot. boost</th>
<th>Wales boost</th>
<th>N. Irel boost</th>
<th>Total sample</th>
<th>Tot adults interviewed</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: 1991</td>
<td>13,840</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13,840</td>
<td>10,264</td>
</tr>
<tr>
<td>B: 1992</td>
<td>12,567</td>
<td>584</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13,151</td>
<td>9,845</td>
</tr>
<tr>
<td>C: 1993</td>
<td>12,219</td>
<td>885</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>13,104</td>
<td>9,600</td>
</tr>
<tr>
<td>D: 1994</td>
<td>11,821</td>
<td>1,030</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12,851</td>
<td>9,481</td>
</tr>
<tr>
<td>E: 1995</td>
<td>11,425</td>
<td>1,124</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12,549</td>
<td>9,249</td>
</tr>
<tr>
<td>F: 1996</td>
<td>11,412</td>
<td>1,308</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12,720</td>
<td>9,438</td>
</tr>
<tr>
<td>G: 1997</td>
<td>11,251</td>
<td>1,301</td>
<td>2,490</td>
<td></td>
<td></td>
<td></td>
<td>15,042</td>
<td>11,193</td>
</tr>
<tr>
<td>H: 1998</td>
<td>11,161</td>
<td>1,300</td>
<td>2,374</td>
<td></td>
<td></td>
<td></td>
<td>14,835</td>
<td>10,906</td>
</tr>
<tr>
<td>I: 1999</td>
<td>10,997</td>
<td>1,339</td>
<td>2,258</td>
<td>3,395</td>
<td>3,577</td>
<td></td>
<td>21,566</td>
<td>15,623</td>
</tr>
<tr>
<td>J: 2000</td>
<td>10,773</td>
<td>1,481</td>
<td>2,193</td>
<td>3,582</td>
<td>3,573</td>
<td></td>
<td>21,602</td>
<td>15,603</td>
</tr>
<tr>
<td>K: 2001</td>
<td>10,624</td>
<td>1,610</td>
<td>2,125</td>
<td>3,516</td>
<td>3,523</td>
<td>5,188</td>
<td>26,586</td>
<td>18,867</td>
</tr>
<tr>
<td>L: 2002</td>
<td>10,470</td>
<td>1,664</td>
<td></td>
<td>3,327</td>
<td>3,385</td>
<td>4,589</td>
<td>23,435</td>
<td>16,597</td>
</tr>
<tr>
<td>M: 2003</td>
<td>10,173</td>
<td>1,701</td>
<td></td>
<td>3,177</td>
<td>3,313</td>
<td>4,210</td>
<td>22,574</td>
<td>16,238</td>
</tr>
<tr>
<td>O: 2005</td>
<td>9,863</td>
<td>1,837</td>
<td></td>
<td>2,985</td>
<td>3,236</td>
<td>3,809</td>
<td>21,730</td>
<td>15,627</td>
</tr>
<tr>
<td>Regions</td>
<td>PSU₁</td>
<td>PSU₂</td>
<td>PSU₃</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wave 1</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wave 2</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wave 3</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Interviewers:**

W 1, 3: \( \text{Interviewer}_1 \), \( \text{Interviewer}_2 \), \( \text{Interviewer}_3 \)

W 2 only: \( \text{Interviewer}_2 \), \( \text{Interviewer}_3 \), \( \text{Interviewer}_1 \)
Complex clustering in the BHPS

- In a panel framework,

\[ Y_{ijkl} = \{\text{micro-level datum}\} \]

<table>
<thead>
<tr>
<th>t = time point</th>
<th>Annual interview, normally September-December</th>
</tr>
</thead>
<tbody>
<tr>
<td>i = individual sample member</td>
<td>OSM / TSM; identified by ‘pid’ (time constant - or ‘cross wave’)</td>
</tr>
<tr>
<td>j = surrounding ‘person group’</td>
<td>Varies by year; ‘person group’ ~ household; identified by ‘hid’ (wave specific identifier)</td>
</tr>
<tr>
<td>k = regional sampling design</td>
<td>Various regional data; we use ‘psu’ = ‘primary sampling unit’ (districts c50k)</td>
</tr>
<tr>
<td>l= interviewer</td>
<td>Usually overlaps regions</td>
</tr>
</tbody>
</table>
BHPS data sources

• Individual and household level data files include identifiers for region clusters, etc
  • *Although restricted access to some identifiers due to the potential risk of identification*

• Wealth of data on relationships between individuals is available from annual ‘egoalt’ data files
  • *This mostly applies to related individuals within the same household*
**‘wEGOALT’ files**

*records are pairs of individuals in the same household, and the relationship between them in a specific wave*

<table>
<thead>
<tr>
<th>bhid</th>
<th>pid</th>
<th>bopid</th>
<th>bsex</th>
<th>bosex</th>
<th>brel</th>
</tr>
</thead>
<tbody>
<tr>
<td>125</td>
<td>2001179</td>
<td>20012179</td>
<td>20012195</td>
<td>1. male</td>
<td>2. lawful spouse</td>
</tr>
<tr>
<td>126</td>
<td>2001179</td>
<td>20012195</td>
<td>20012179</td>
<td>2. female</td>
<td>2. lawful spouse</td>
</tr>
<tr>
<td>127</td>
<td>2001195</td>
<td>20012519</td>
<td>20012535</td>
<td>2. female</td>
<td>2. female</td>
</tr>
<tr>
<td>128</td>
<td>2001195</td>
<td>20012535</td>
<td>20012519</td>
<td>2. female</td>
<td>4. natural child</td>
</tr>
<tr>
<td>129</td>
<td>2001217</td>
<td>10023526</td>
<td>10023569</td>
<td>2. female</td>
<td>13. natural parent</td>
</tr>
<tr>
<td>130</td>
<td>2001217</td>
<td>10023526</td>
<td>20011008</td>
<td>2. female</td>
<td>10. natural brother/</td>
</tr>
<tr>
<td>131</td>
<td>2001217</td>
<td>10023569</td>
<td>20011008</td>
<td>1. male</td>
<td>2. female</td>
</tr>
<tr>
<td>132</td>
<td>2001217</td>
<td>10023569</td>
<td>20011008</td>
<td>2. female</td>
<td>2. female</td>
</tr>
<tr>
<td>133</td>
<td>2001217</td>
<td>20011008</td>
<td>10023526</td>
<td>1. male</td>
<td>10. natural brother/</td>
</tr>
<tr>
<td>134</td>
<td>2001217</td>
<td>20011008</td>
<td>10023569</td>
<td>2. female</td>
<td>12. brother/sister-</td>
</tr>
<tr>
<td>135</td>
<td>2001233</td>
<td>20012888</td>
<td>20012918</td>
<td>2. female</td>
<td>2. female</td>
</tr>
<tr>
<td>136</td>
<td>2001233</td>
<td>20012918</td>
<td>20012888</td>
<td>2. male</td>
<td>13. natural parent</td>
</tr>
<tr>
<td>137</td>
<td>2001306</td>
<td>20013507</td>
<td>20013523</td>
<td>1. male</td>
<td>23. lodger/border</td>
</tr>
<tr>
<td>138</td>
<td>2001306</td>
<td>20013507</td>
<td>20013558</td>
<td>1. male</td>
<td>23. lodger/border</td>
</tr>
<tr>
<td>139</td>
<td>2001306</td>
<td>20013507</td>
<td>20013574</td>
<td>1. male</td>
<td>23. lodger/border</td>
</tr>
<tr>
<td>140</td>
<td>2001306</td>
<td>20013523</td>
<td>20013574</td>
<td>2. female</td>
<td>27. landlady/lord</td>
</tr>
<tr>
<td>141</td>
<td>2001306</td>
<td>20013523</td>
<td>20013574</td>
<td>2. female</td>
<td>4. natural child</td>
</tr>
<tr>
<td>142</td>
<td>2001306</td>
<td>20013523</td>
<td>20013558</td>
<td>1. male</td>
<td>4. natural child</td>
</tr>
<tr>
<td>143</td>
<td>2001306</td>
<td>20013558</td>
<td>20013523</td>
<td>1. male</td>
<td>13. natural parent</td>
</tr>
<tr>
<td>144</td>
<td>2001306</td>
<td>20013558</td>
<td>20013507</td>
<td>1. male</td>
<td>27. landlady/lord</td>
</tr>
<tr>
<td>145</td>
<td>2001306</td>
<td>20013558</td>
<td>20013574</td>
<td>1. male</td>
<td>10. natural brother/</td>
</tr>
<tr>
<td>146</td>
<td>2001306</td>
<td>20013574</td>
<td>20013558</td>
<td>2. female</td>
<td>10. natural brother/</td>
</tr>
<tr>
<td>147</td>
<td>2001306</td>
<td>20013574</td>
<td>20013507</td>
<td>2. female</td>
<td>10. natural brother/</td>
</tr>
<tr>
<td>148</td>
<td>2001306</td>
<td>20013574</td>
<td>20013523</td>
<td>2. female</td>
<td>13. natural parent</td>
</tr>
</tbody>
</table>
Attention to clustering in the BHPS

• **O’Muircheartaigh and Campanelli** found small but significant regional and interviewer effects, e.g.

• **Johnston et al. (2005)** explored household context of voting and noted substantial empirical clustering effects

• **Chandola et al. (2003)** explored household context of subjective health and noted strong household influences

• **The most common approach is to ignore clusters.**
  - Individual level analyses (sometimes with individual level weights)
    - Occasional use of models with extra control for shared variance, e.g. ‘robust clusters’
    - Some analyses remove household clusters de facto (e.g. men only)
  - Household level context
    - Individual level models with mix of individual level and household level variables
    - Universal application of the common UK definition of household
In this paper...

1) Defining / exploring the person group context
   - Different types of ‘person group’ (cf. household)
   - Longitudinal treatments for ‘person groups’

2) Alternative modelling strategies
   - Multilevel / random effects
   - Other regression approaches

Pragmatic conclusions
1) Some possible ‘person groups’ (PGPs)

<table>
<thead>
<tr>
<th>PGP/HH ID's/ PGP</th>
<th>BHPS Wave 2 (1992)</th>
<th>ID’s/ PGP</th>
<th>PGP/HH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual</td>
<td>ID</td>
<td>Single people only</td>
<td>1.00; 1.00</td>
</tr>
<tr>
<td>Couple</td>
<td>CP</td>
<td>Cohabiting couples</td>
<td>1.44; 1.33</td>
</tr>
<tr>
<td>Minimal</td>
<td>MH</td>
<td>Couple or single parent plus any dependent children</td>
<td>1.47; 1.80</td>
</tr>
<tr>
<td>Household Unit</td>
<td>FA</td>
<td>Couple or SP plus unmarried children; grandparent-child if carer</td>
<td>1.69; 2.08</td>
</tr>
<tr>
<td>(Inner) Family</td>
<td>CU</td>
<td>All household sharers related by blood, marriage or guardianship</td>
<td>1.80; 2.39</td>
</tr>
<tr>
<td>Household</td>
<td>HH</td>
<td>All living in same building who share meals or living room</td>
<td>1.88; 2.52</td>
</tr>
<tr>
<td>All waves</td>
<td>XH</td>
<td>All living in any HH’s to have shared ID’s in any previous wave</td>
<td>1.96; 2.61</td>
</tr>
<tr>
<td>Category</td>
<td>BHPS Wave 15 (2005)</td>
<td>ID’s/ PGP</td>
<td>PGP/HH</td>
</tr>
<tr>
<td>----------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>----------------</td>
<td>---------------</td>
</tr>
<tr>
<td><strong>Individual</strong></td>
<td><strong>ID</strong> Single people only</td>
<td>1.00; 1.00</td>
<td>1.80; 2.50</td>
</tr>
<tr>
<td><strong>Couple</strong></td>
<td><strong>CP</strong> Cohabiting couples</td>
<td>1.41; 1.32</td>
<td>1.27; 1.88</td>
</tr>
<tr>
<td><strong>Minimal Household Unit</strong></td>
<td><strong>MH</strong> Couple or single parent plus any dependent children</td>
<td>1.44; 1.45</td>
<td>1.25; 1.72</td>
</tr>
<tr>
<td><strong>(Inner) Family</strong></td>
<td><strong>FA</strong> Couple or SP plus unmarried children; grandparent-child if carer</td>
<td>1.56; 1.56</td>
<td>1.15; 1.60</td>
</tr>
<tr>
<td><strong>Consumer Unit</strong></td>
<td><strong>CU</strong> All household sharers related by blood, marriage or guardianship</td>
<td>1.75; 2.40</td>
<td>1.02; 1.04</td>
</tr>
<tr>
<td><strong>Household</strong></td>
<td><strong>HH</strong> All living in same building who share meals or living room</td>
<td>1.80; 2.50</td>
<td>1.00; 1.00</td>
</tr>
<tr>
<td><strong>All waves Household</strong></td>
<td><strong>XH</strong> All living in any HH’s to have shared ID’s in any previous wave</td>
<td>2.17; 2.93</td>
<td>0.85; 0.83</td>
</tr>
</tbody>
</table>
Person Group Sizes, BHPS Wave 2 enumerated sample
(Excluding 5 hhlds with 10+)

Lambert/Gayle, RC33 2008
Calculating ‘person group’ identifiers?

- A sequence of operations on one ID’s eligibility to be in another ID’s PGP
- Aggregated within waves to individual level file
- Stata> do http://www.longitudinal.stir.ac.uk/bhps/bhps_1to15_pgp.do
Longitudinal analysis & wave-specific PGPs?

- Tractable solutions
  - ‘All wave PGP’ = *at any given wave, a cluster defined by all pids in the wave who are now, or have every been, in the same household/pgp at any point in the preceding survey*
    - Easily defined (see above ‘XH’ for households)
    - Groups expand in size over survey waves
    - Realistic way to recognise inter-respondent connections *in cross-sectional analysis*
    - Can support an additional nested cluster for the current PGP
  - ‘Longitudinal PGP’ = *For a random pid within the PGP at a chosen wave, all pids who are in the same PGP at any other point in time*
    - Simple nested model amenable to panel data analysis
    - Rejects cases outside the pgp, and ignores other possible PGPs

- Models for ‘non-nested’ structures
  - ‘Cross classified’ / ‘multiple membership’ models
    - Feasible, but computationally demanding and may be subject to identification problems
### Example: longitudinal households

<table>
<thead>
<tr>
<th>Household</th>
<th>BHPS Wave 15 (2005)</th>
<th>ID’s/ PGP</th>
<th>PGP/HH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Adult intrv.; ennumerated</td>
<td></td>
</tr>
<tr>
<td>Household</td>
<td>HH</td>
<td>1.80; 2.50</td>
<td>1.00; 1.00</td>
</tr>
<tr>
<td></td>
<td><em>Within a wave, all living in same building who share meals or living room</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All waves household</td>
<td>XH</td>
<td>2.17; 2.93</td>
<td>0.85; 0.83</td>
</tr>
<tr>
<td></td>
<td><em>All living in any HH’s to have shared ID’s in any previous wave</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Longitudinal Household</td>
<td>LH</td>
<td>1.80; 2.50</td>
<td>1.00; 1.00</td>
</tr>
<tr>
<td></td>
<td><em>For one selected individual, all indiv’s who currently share the HH (for w15)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LH</td>
<td><strong>16.4</strong></td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td><em>(min 1, max 61)</em></td>
<td>(= 1/15)</td>
<td></td>
</tr>
</tbody>
</table>
2) Assessing the impact of PGP patterns

• Relative size of variance components

• Impact of hierarchical structures upon regression model coefficients
  – Similarity and efficiency
  – Dependence and bias
Person Group level variance components for selected models

Source: BHPS 1992, random effects in Stata with xtreg / xtlogit
Person Group level variance components for selected models

Null CAMSIS Reg (CAMSIS ~ educ., social background)
Null GHQ Reg (GHQ ~ social circ., partner GHQ)
Null DD Logit (Degree/Dip ~ social background)
Null ConVote Logit (ConVote ~ social circ.
Logit (ConVote ~ social circ., partner voting)

Source: BHPS 1992, random effects in Stata with xtreg / xtlogit
**Significant deviance reductions: modelling person groups variance components within gender groups**

(Null models on cross-sectional data wave 2, for indvs within PGP’s within PSU regions; from Lambert 2001)

<table>
<thead>
<tr>
<th></th>
<th>Men only</th>
<th>Women only</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mu</td>
<td>Fa</td>
</tr>
<tr>
<td>Personal income</td>
<td>nc</td>
<td>nc</td>
</tr>
<tr>
<td>Wage income</td>
<td>nc</td>
<td>nc</td>
</tr>
<tr>
<td>CAMSIS</td>
<td>o-</td>
<td>+</td>
</tr>
<tr>
<td>Occ advantage</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Subjective class</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Degree/Diploma</td>
<td>+</td>
<td>o-</td>
</tr>
</tbody>
</table>

{blank} : no; ● yes; ○ marginal; nc : convergence not achieved - usually reflects non-significant VC)
Example: Predicting CAMSIS score for current job, wave B, for cohabiting working adults aged 30-60

<table>
<thead>
<tr>
<th></th>
<th>Linear regression</th>
<th>3-level random effect</th>
<th>Linear regression</th>
<th>Linear reg. + Heckman select</th>
<th>3-level after Heckman sel.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Indv; CP; PSU</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fath CAMSIS</td>
<td>0.23 (0.02)</td>
<td>0.21 (0.02)</td>
<td>0.17 (0.02)</td>
<td>0.18 (0.02)</td>
<td>0.17 (0.02)</td>
</tr>
<tr>
<td>Deg/Diploma</td>
<td>10.9 (0.5)</td>
<td>10.4 (0.5)</td>
<td>9.0 (0.5)</td>
<td>9.7 (0.6)</td>
<td>9.9 (0.6)</td>
</tr>
<tr>
<td>Blck. Carib.</td>
<td>-11.5 (4.0)</td>
<td>-11.7 (4.2)</td>
<td>-9.0 (3.8)</td>
<td>-9.7 (3.8)</td>
<td>-9.8 (3.8)</td>
</tr>
<tr>
<td>Blck. Oth.</td>
<td>2.3 (4.5)</td>
<td>2.4 (4.6)</td>
<td>0.2 (4.3)</td>
<td>0.1 (4.3)</td>
<td>0.3 (4.3)</td>
</tr>
<tr>
<td>Indn.</td>
<td>-6.5 (2.1)</td>
<td>-6.9 (2.2)</td>
<td>-5.0 (2.0)</td>
<td>-4.8 (2.0)</td>
<td>-5.0 (2.0)</td>
</tr>
<tr>
<td>Sp. CAMSIS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.24 (0.02)</td>
<td>0.24 (0.02)</td>
<td>0.23 (0.02)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sp. Deg/Dip</td>
<td></td>
<td></td>
<td>1.6 (0.6)</td>
<td>1.6 (0.6)</td>
<td>1.7 (0.6)</td>
</tr>
<tr>
<td>Lambda</td>
<td></td>
<td></td>
<td></td>
<td>4.8 (2.0)</td>
<td>-9.1 (4.3)</td>
</tr>
<tr>
<td>PGP VC</td>
<td>5.3 (0.5)</td>
<td></td>
<td></td>
<td></td>
<td>1.3 (0.5)</td>
</tr>
</tbody>
</table>
Example: Predicting GHQ (good subjective well-being) for adults in wave 15 using ‘all wave person groups’ (*HH level*, *‘Essex’ sample*)

<table>
<thead>
<tr>
<th></th>
<th>Linear regression</th>
<th>Linear reg. + HW robust (XH)</th>
<th>Random effects (xtmixed in Stata)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$a$</td>
<td>$b$</td>
<td>$a$</td>
</tr>
<tr>
<td>Female</td>
<td>-0.5**</td>
<td>-0.6**</td>
<td>-0.5**</td>
</tr>
<tr>
<td>Cohabitng</td>
<td>0.3**</td>
<td>0.3**</td>
<td>0.3**</td>
</tr>
<tr>
<td>Age</td>
<td>$U^*$</td>
<td>$U^*$</td>
<td>$U^*$</td>
</tr>
<tr>
<td>Own CAMSIS</td>
<td>0.9**</td>
<td>0.8**</td>
<td>0.9**</td>
</tr>
<tr>
<td>Sp. –GHQ</td>
<td>-0.1**</td>
<td>-0.1**</td>
<td>-0.1**</td>
</tr>
<tr>
<td>VC at PSU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VC at XH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VC at CP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VC at ID</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Lambert/Gayle, RC33 2008
Contribution of fixed effects estimators for within PGP change?

Repeated Measures Data

Comparison of Beta estimates in three models

- Observed
- ols (clustered)
- R.E.
- F.E. (overall)
- individual 1
- individual 2
- individual 3
- individual 4

Vernon Gayle & Paul Lambert
Example: Predicting CAMSIS score for current job, wave B, for cohabiting working adults aged 30-60

<table>
<thead>
<tr>
<th></th>
<th>Linear regression</th>
<th>Lin Reg. robust cluster</th>
<th>Pop. Average (GEE)</th>
<th>Random effects</th>
<th>Fixed effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>2427 adults within 1634 person groups (CP – couples)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fath CAMSIS</td>
<td>0.23**</td>
<td>0.23**</td>
<td>0.22**</td>
<td>0.21**</td>
<td>0.11**</td>
</tr>
<tr>
<td>Deg/Diploma</td>
<td>10.9**</td>
<td>10.9**</td>
<td>10.5**</td>
<td>10.4**</td>
<td>6.9**</td>
</tr>
<tr>
<td>Blck. Carib.</td>
<td>-11.5*</td>
<td>-11.5*</td>
<td>-11.7*</td>
<td>-11.7*</td>
<td>-8.5</td>
</tr>
<tr>
<td>Blck. Oth.</td>
<td>2.3</td>
<td>2.3</td>
<td>2.6</td>
<td>2.6</td>
<td>9.9</td>
</tr>
<tr>
<td>Indn.</td>
<td>-6.5*</td>
<td>-6.5*</td>
<td>-6.4*</td>
<td>-6.3*</td>
<td>-7.5</td>
</tr>
<tr>
<td>P-value of test BlckC.≠BlckO.</td>
<td>0.02</td>
<td>0.06</td>
<td>0.02</td>
<td>0.02</td>
<td>0.29</td>
</tr>
</tbody>
</table>
Example: Predicting conservative voting preference in panel analysis for adults in waves 1-15, with and without LH clustering patterns

<table>
<thead>
<tr>
<th></th>
<th>Logit regression</th>
<th>Random effects panel (Sabre)</th>
<th>Random effects panel plus PGP at LH level (Sabre)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
<td>b</td>
<td>a</td>
</tr>
<tr>
<td>Female</td>
<td>0.10**</td>
<td>0.08**</td>
<td>0.10**</td>
</tr>
<tr>
<td>Age</td>
<td>0.02</td>
<td>0.01**</td>
<td>0.03**</td>
</tr>
<tr>
<td>Wave*10</td>
<td>-1.0</td>
<td>-0.35**</td>
<td>-1.0**</td>
</tr>
<tr>
<td>GHQ.*10</td>
<td>0.26**</td>
<td>0.17**</td>
<td>0.26**</td>
</tr>
<tr>
<td>Lag Convot.</td>
<td>4.51**</td>
<td></td>
<td>4.57**</td>
</tr>
<tr>
<td>VC at LH</td>
<td></td>
<td></td>
<td>2.42**</td>
</tr>
<tr>
<td>VC at ID</td>
<td>0.40**</td>
<td>0.31**</td>
<td>2.86**</td>
</tr>
<tr>
<td>VC at t</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Random effects panel plus PGP at LH level (Sabre)
a: n=23874; 132755 units; 8657 LHs
b: n=21433; 114528 units; 8464 LHs

Lambert/Gayle, RC33 2008
Pragmatic conclusions

1) **Person group clustering as ‘similarity’** can largely be ignored

- **PGP effects are significant but of negligible consequence**
  - Different types of PGP seldom matter (except for some processes)
  - Clustering component is most likely to impact effect of skewed variables
  - Reducing analysis to male/female only is a robust option

*Panel analysis:*
- Cross-wave PGP clusters (‘XH’) are little different to household based clusters
- Software considered here:
  - SabreStata a convenient estimator for up to 3 level nested models
  - Stata (xtmixed)
  - MLwiN

2) **Person group clustering as ‘dependence’** may matter much more

- Substantial effects of predictors derived from the person group
  - Fixed effects estimators and other model specifications (e.g. random effects with random coefficients) can be used to give alternative emphases

*Panel analysis*
- contribution of variable constructions for other household sharers
References


