

ADMIN – Administrative Data: Methods, Inference and Network

Lorraine Dearden

Director ADMIN

Institute of Education

Email: l.dearden@ioe.ac.uk

The team

- Quantitative Social Scientists from a number of disciplines including Social Statistics, Economics and Sociology
- CIs are
 - Mac McDonald
 - Anna Vignoles
 - Nikos Tzavidis
 - Sophia Rabe-Hesketh
 - James Brown
 - Kirstine Hansen
- Based in DQSS and CLS at IoE

Research Programme

- Involves exploiting new and forthcoming linked administrative and survey data to undertake 2 programmes of research to:
 1. Establish how best to use administrative data (4 projects)
 2. How we can enhance analysis of existing longitudinal survey data (4 projects)

1. Improving analysis of Admin Data

- Administrative data is widely used for policy making and to measure effectiveness of public service provision
 - E.g school and hospital league tables, evaluating policies
- BUT administrative data typically doesn't have extensive background information on students, patient etc which also impact on outcome of interest
 - Largely untested whether methodological approaches taken actually measures question of interest

Example 1: CVA

- Contextualized value added (CVA) is hot policy topic
- Recognised the fact that traditional league tables did not take into account differences in prior academic achievement or socio-economic background of students
 - Disadvantaged schools in poor areas did poorly even if adding a huge amount of value added by school to the child's education
- CVA seen as solution to this

Can test this with linked data....

- For group of children born in 1989/90 we have a large longitudinal survey (15,000) which has been linked to NPD
- Next Steps has much richer background characteristics such as family income, parental education, family structure and composition etc
- Can test using this linked data whether CVA does what it is meant to – measure value added by school

How?

- CVA can only control for variables in NPD data
- What we see in Next Steps, however, is that things like home tutoring, time spent with child doing homework, parental education etc has significant impact on child outcomes
- Can't control for this in NPD so school rather than parents may get credit for this
- Implications for school league tables
 - Depends crucially whether this behaviour is a supplement or complement to the job being done by the school and methodological approach used to calculate CVA
- With Next Steps linked to NPD can come up with methods to limit these types of biases in CVA measures using just admin data

Example 2: Can NPD data proxy social disadvantage?

- Huge debate about how good proxies in school administrative data are for social disadvantage
- Have good ethnicity measures, fsm status, but no information on parents employment, education or income/wealth
- Generally use postcode/local area information to proxy these missing covariates
- But how do they perform?
- Can assess this using NPD linked to Next Steps

Is FSM a good proxy of disadvantage?

- Construct a measure of socio-economic position using data on income, and occupation of parents
- This is what FSM should capture as only eligible if on benefits
- Also test whether we can come up with better measures using just admin data

OK but not great.....

| Quintile of SEP | Not on FSM (%) | On FSM (%) |
|-----------------|----------------|------------|
| 1 (Bottom) | 9.78 | 10.21 |
| 2 | 17.32 | 2.66 |
| 3 | 19.39 | 0.61 |
| 4 | 19.70 | 0.31 |
| 5 (Top) | 19.96 | 0.05 |
| TOTAL | 86.16 | 13.84 |

2. Improving Analysis of Longitudinal Data

- Possibility of new ways of dealing with perennial problems associated with longitudinal data e.g. attrition bias, recall bias...
 - E.g we know exactly how well children who drop out of Next Steps perform at school as we can follow them in the administrative data
- Admin data may be used to fill in gaps in survey data and vice versa
 - E.g. in MCS asking for DWP linkage for both parents so if families split up can still cover labour market outcomes of absent father....

ADMIN: Training and Capacity Building

- Extensive range of courses focussing on Methodological issues around data linkage
 - Why data linkage is important
 - How best to use this new data linkage (several courses on methods for different problems)
 - How to actually undertake data linkage

Conclusions

- More and more survey data sets are going to be linked to admin data and this has huge potential
- ADMIN will be a crucial resource in exploiting this new resource but also in developing researcher capacity in using this new and exciting source of data