Harry B.G. Ganzeboom

The Comparative Measurement of Stratification Indicators

Mini-Course, University of Tallinn
November 17-19 2010

Lecture Day 3

Social Class Measures and Models

9. The third generation

History

- Around 1975 everything was ready for replicating the BD model: researchers, data, data-archives, computers, SPSS, LISREL.
- In 1975 American sociologist Robert Hauser (working on the OCG replication) imported log-linear modeling into sociology.
- At about the same time British sociologist John Goldthorpe (working on the Oxford Mobility Inquiry) changed his mind in the same direction.
- Their books: “Opportunity and Change” (1978), “Social Mobility and Class Structure in Britain” (1980).

EGP

- In 1979 Goldthorpe teamed up with Swedish sociologist (with co-author Lucienne Portocarero) to produce a three-country comparison (UK-FR-SE) using 'topological' log-linear models.
- This was the start of the CASMIN project, that culminated in "The Constant Flux" (1992) – comparing 12 countries.
- The 1979 article introduced the EGP classification into the field.

EGP schema (1)

- In his British work Goldthorpe had worked with a 7-category scheme:
  - I Higher Managers & Professionals
  - II Lower Managers & Professionals
  - III Routine Clerical and Sales Workers
  - IV Small Proprietors
  - V Manual Supervisors
  - VI Skilled Manual Workers
  - VII Semi- and Unskilled Manual Workers
- Notice that farmers have no distinct category!!

EGP schema (2)

- For comparative work, it soon became clear that you have to separate out farm work (IV-c, VII-b).
- Another distinction was between IV-a and IV-b (small proprietors with and without employees).
- Later refinement: split up III into III-a (mostly Clerical) and III-b (mostly Service & Sales).
- Initially:
  - No theoretical justification was provided;
  - No general way to code EGP from data was published.

The theory on EGP

- See Rose & Harrison: “Work and market relationships”.
- Basic positions are:
  - Employers
  - ‘Service’ relationship
  - Self-employment
  - Labour relationship
- These positions are generated by (A) Monitoring, and (B) Specificity of Human Assets (= required skills).

ISCO and EGP

- Untill well after 1992 there was no documentation how EGP was coded from data.
- In 1984 I visited Goldthorpe to find out how an EGP schema for the Netherlands was to be constructed and I used ISCO-68 to translate the NL census classification into English.
- After some communication this generated an EGP algorithm coded from the NL classification, but for ISCO-68 at the same time!
- First published in 1989, together with the application on social mobility data from 35 countries (Ganzeboom, Luijkx & Treiman, 1989).

ISCO and EGP

- In 1995 the CASMIN recoded were made available, but they were hard to apply in cross-national research.
- The ISCO → EGP algorithm was upgraded for ISCO-88 and published in1996. This has been heavily used by the research community.

EGP [HG version]

- Self-employment: (1) Salaried, (2) Self-employed.
- Supervising status: (0) None (5) 1-10 (11) 11+ (25) 25+.
- Occupational sector: Farm, Manual, Non-manual.
- Skill: Lower, Medium and Higher.
- This implies 2*3*3*3 = 54 combinations. These are condensed into 10.
- The algorithm starts at occupations. Then the initial assignment is refined by taking into account self-employment and supervising status.
- If supervising status is missing in data, the scheme simplifies. If self-employment is missing, some classes (smaller self-employed) cannot be distinguished.

ESEC

- More recently British sociologist David Rose had created an ESEC scheme which is much inspired by EGP.
- As a matter of fact it is a straight recode of ISCO-88 into a scaled-down version of EGP.
- It started in Britain and then became international. Note that farm workers are again forgotten – a British peculiarity.

10. Methods of social mobility analysis

Correlations

- The BD model is built upon correlation (linear association = covariance) and regression as statistical models.
- These types of coefficients are simple, but they are sensitive to changes in marginal distributions.
- This is a problem, as marginal distributions in comparative mobility research tend to be different between countries, and between time periods.
- I can simply demonstrate this in SPSS.
- There are also association measures that are insensitive to marginal multiplication: odds-ratio and derived association measures (Yule's Q and Goodman's Gamma).

Log-linear odds-ratio models

- Mobility tables can be regarded as a bunch of (contiguous) odds-ratio's, with the special feature that row and column represent the same variables ('square tables').
- Log-linear models can be regarded as constraints (simplifications) on all odd-ratio's in a table.
- E.g. a very simple model would specify that all contiguous odds ratio's are the same ('uniform'). This model is called the uniform association model.
- In a UA model all the association is represented in one number, much like a regression or correlation coefficient. In one special condition (normal distribution of the marginals) UA is even the same as a pearson correlation.

Quasi-uniform

- UA models often do not fit the data. In occupation mobility tables it is common to find higher densities on the diagonal of a tables (immobility).
- In many models these diagonal 'effects' receive separate treatment. These type of models are often referred to a Quasi-Uniform models.

Scaled association models

- Goodman (1972) developed another refinement to the UA model: scaled association.
- In these models odds-ratio's are constrained to be dependent upon distances between categories. These distances (scalings) can be estimated from the data, or be introduced as fixed (e.g. ISEI can be used for this).
- RC-II models estimate the scalings from the data: they turn nominal classifications into distances!

Topological models

- Not everybody likes the idea of using distance to model social mobility. Many researchers (following Erikson & Goldthorpe) model association patterns by singling out specific combination of cells with high of low density ('topological' models).
- While topological models still are insensitive to marginal distributions, they do not produce simple odds-ratio's and they tend to be complicated by many parameters.
- The CASMIN 'core' model is of this kind.

Plusses and minuses

+ Loglinear models have brought more refined insights into mobility patterns.
+ Marginal distribution do not change uniformly, densities do not follow normal distribution, and likely wrong comparative conclusions have been drawn using correlational methods.
- However, it is all about bivariate distributions.
- Measurement error has been totally overlooked.

Multinomial logit models

- Log-linear modeling is closely related to the multinomial logit model, which can be applied to individual data and allows for covariates (multivariate analysis).
- Multinomial models are now often applied (in particular in the education – occupation link), but have three related disadvantages:
  - They produce an enormous amount of parameters.
  - They are not informed by metric consideration.
  - They have low statistical power.

Conditional multinomial logit models

- McFadden (1977, Nobel Prize 2000) has proposed conditional multinomial logit models, that can integrate log-linear models with multinomial models using individual data.
- Hendrickx & Ganzeboom (1998) and Dessens et al. (2003) explain how to do this in a social-mobility context.
- The basic idea is very similar to event history of panel analysis: every respondent is represented in the data multiple times, i.e. as often as s/he faces choices (EGP categories).
- We then run a logistic model on which alternative s/he has actually chosen.
- Scalings, diagonal effects or other topological varieties can be build into the CLR model.

But ..

- While CLR combines the virtues of loglinear analysis and multiple regression…
- … it cannot easily be extended to incorporate models for measurement error.
- Promising alternatives: Latent class models or Latent Markov models can take measurement error into account. We have yet to see how these models work out on stratification data with multiple indicators.