Harry B.G. Ganzeboom

The Comparative Measurement of Stratification Indicators

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

Lecture 1

Organization of the Course

Causal and Measurement Models

History of Stratification Research

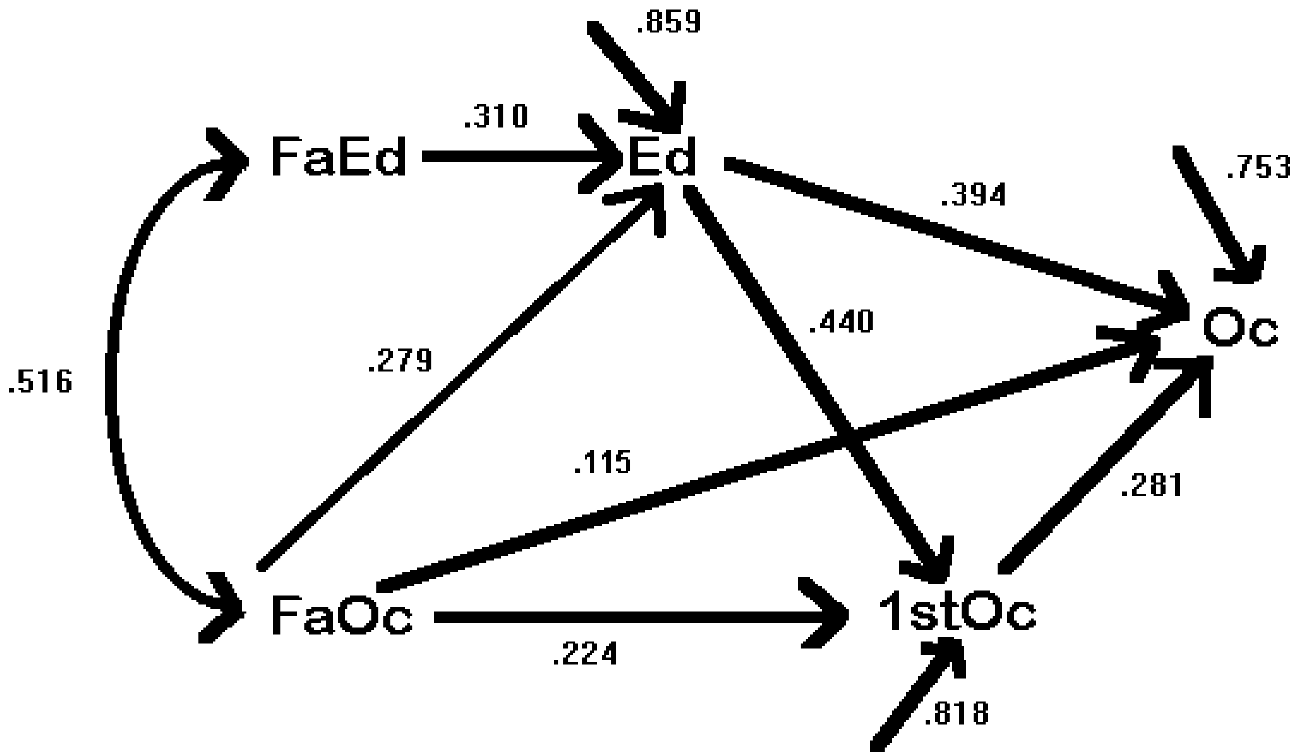
1. Organization of the course

Contents of the Course

- Comparative measurement of Education, Occupation and Income.
- Six lectures
  - Status Attainment Model
  - Measuring Occupational Status
  - Occupation Coding
  - Social Class
  - Education
  - Income
- Readings
- Six questions to be submitted to Harry.Ganzeboom@gmail.com

2. The status attainment model

BD Status attainment model (1967)



Blau & Duncan (1967)

- In 1967 B&D revolutionized research in social stratification (and sociology in general) by the introduction of a "causal" model for the "process of stratification".
- The model decomposed bivariate correlations among causally ordered variables into partial (causal) effects.

Simple correlations USA 1962

B&D Table1: Simple Correlations for Five Status Variables					
	FaEd	FaOc	Ed	1stOc	Occ
FaEd	1.000				
FaOc	0.516	1.000			
Ed	0.453	0.438	1.000		
1stOc	0.332	0.417	0.538	1.000	
Occ	0.322	0.405	0.596	0.541	1.000

Causal Effects

- Causal effects are usually established in controlled randomized experiments and specify how consequence Y changes due to manipulation of cause X.
- Effects in a path model can be interpreted as causal, provided that:
  - The variables are causally ordered (no reverse causation).
  - The control variables are completely specified.
- Note that these two conditions together replace the randomized design of a controlled experiment.

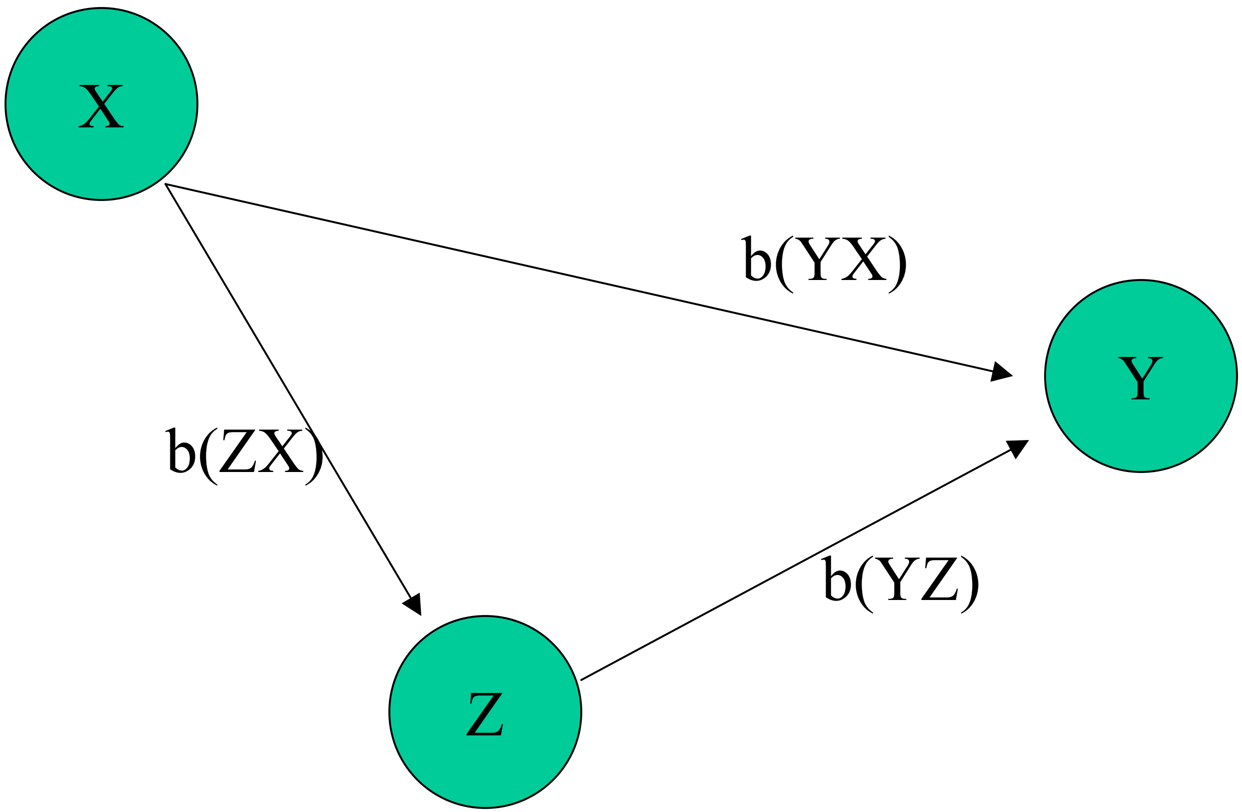
Direct and indirect effect

- Indirect effects occur when X influences Y via an intervening / mediation variable Z (causal chain).
- X → Z → Y
- Size of indirect effect is identical to the multiplication of the two component direct effects.
- Z can be thought of as an explanation how / why the effect of X on Z arises.
- Note that the indirect effects still imply that X influences Y.

Confounding effects and spurious association

- A confounding effect arises when a prior variable Z influences both X and Y and creates association between X and Y.
- Y ← Z → X
- Confounding (spurious) effects are also (!) equal to the multiplication of the two direct effects.
- Although Z can be thought of as an explanation of how / why association between X and Z arises, there is NO causal effect of X on Y.
- It is therefore more appropriate to speak of "spurious association" in stead of "spurious effects".

Elementary causal model



The elementary algebra

- When assuming a standardized linear process (correlation, regression), it is easy to calculate direct and indirect effects in the elementary causal model.
- Basic equation of causal path analysis:

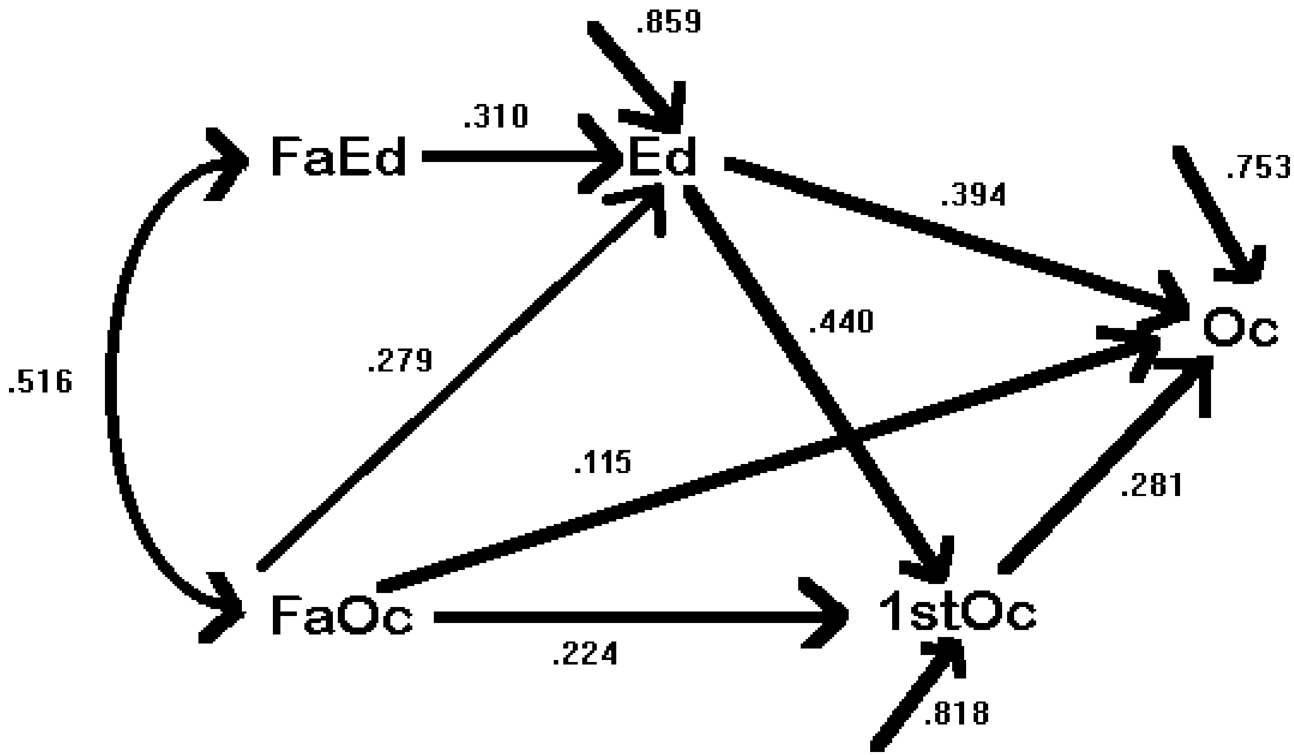
  Total correlation = direct effect + indirect effects + confounding effects.

- All direct effects in the model can be calculated from three equations:
  - bZX = rZX
  - bYX + bZX*bYZ = rYX
  - bYX + bZX*bYX = rYZ

More elementary algebra

- Then we can calculate:
  - Indirect effect: bYX = bYZ*bZX
  - Spurious association: bZX*bYX.
- Note that the basic equation of path analysis is additive, so we can express direct, indirect and confounding effects as percentages of the total correlation.
- We can easily do the calculation by hand, but for more complicated models we use a SEM (simultaneous equation model) program, such as LISREL
- In a fully saturated model (all arrows present), the effects are equal to standardized regression coefficients. In an incomplete model (some arrows omitted) they are **NOT**.

BD Status attainment model (1967)



Conclusions from the BD model

comparisons of causal influences

- Father’s occupation and father’s education are about equally important for determining educational outcomes.
- Father’s occupation contributes to occupational status attainment at entry into the labor market and also after entry into the labor market (father’s education does not).
- Respondent’s education is about (only!) twice as important for entry status as father’s occupation.
- Respondent’s education is about three times as important for final status as father’s occupation.

More conclusions:
indirect effects

- About half of the total social reproduction at labor market entry is explained by selective processes in education, the other half is direct (unexplained).
- Social reproduction in current occupation is slightly stronger than at career’s beginning and a larger share is explained by selective processes in education.

Additional methodological contributions by B&D

- Measuring all variables at an interval scale (necessary for calculating correlation and regression.
- Large scale survey (OCG: Occupational Change in a Generation).
- First job: (A) makes it possible to look at career dynamics, (B) allows for cohort comparisons to assess historical change.

Theoretical contribution

- Modernization theory (from ascription to achievement): “As societies modernize, education becomes more important and father’s occupation becomes less important for occupational attainment.”
- It is tested by Table 5.4 or correlations in Table 5.3): is it confirmed or not?

Quasi full-career model

- In Figure 5.2 the same information is expressed in a ‘synthetic cohort analysis’ (as if they had panel data!).
- Observe
  - that father’s influence is strongest at career beginnings,
  - But that educational influence is still strong after career beginning
  - And that careers stabilize during the life cycle.
- All of this from a simple cross-sectional dataset!

3. Measurement Theory

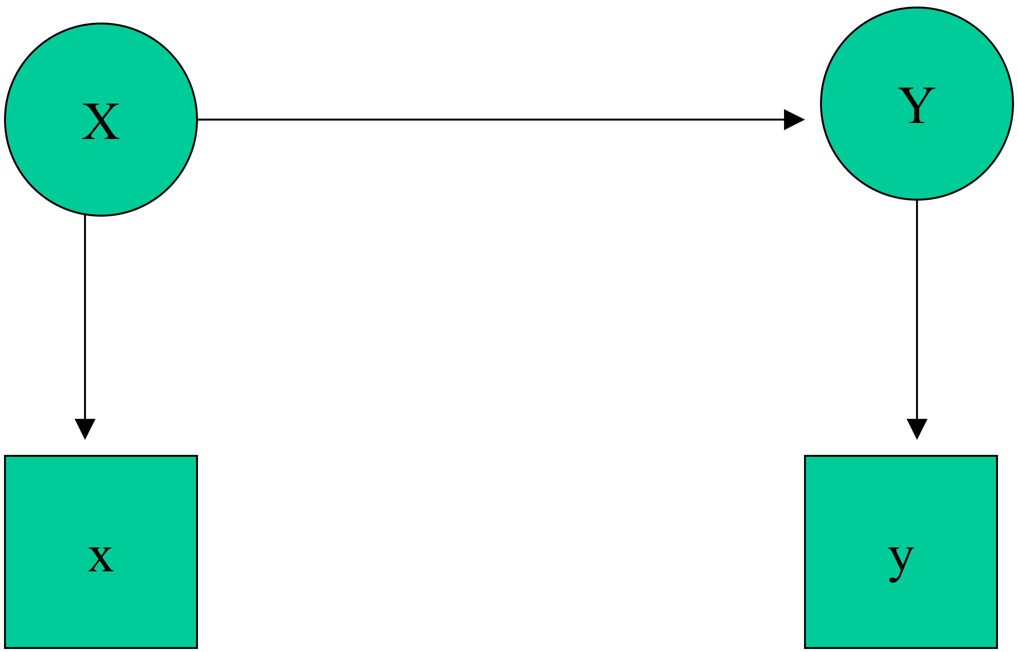
Causal approach to measurement error

- It is useful to think of measurement as a causal process, in which a latent variable <u>causes</u> the observed indicator.
- Latent variable: true score == how the world really is.
- Observed indicator: the way the world ended up in your data-matrix.
- Indicator = true score + measurement error.

Random and systematic error

- Random error arises when the error is totally unpredictable, as if occurring by throwing a dice.
- This implies that random errors are not associated with any other variables, inside or outside the model, including other errors.
- Random error is also called **unreliability**.
- Systematic error (**invalidity**) is different: see further below.

A model with random measurement error



Consequences of (random) measurement error

- Let’s assume that in a bivariate causal relationship the two variables are measured with random error.
- Using the elementary algebra of path analysis we see that such measurement error reduces (‘attenuates’) the observed effect relative to the true effect.
- We can turn this argument around (‘correction for attenuation’): if we would know the measurement relationships, we could estimate the true relationship from the observed relationship.

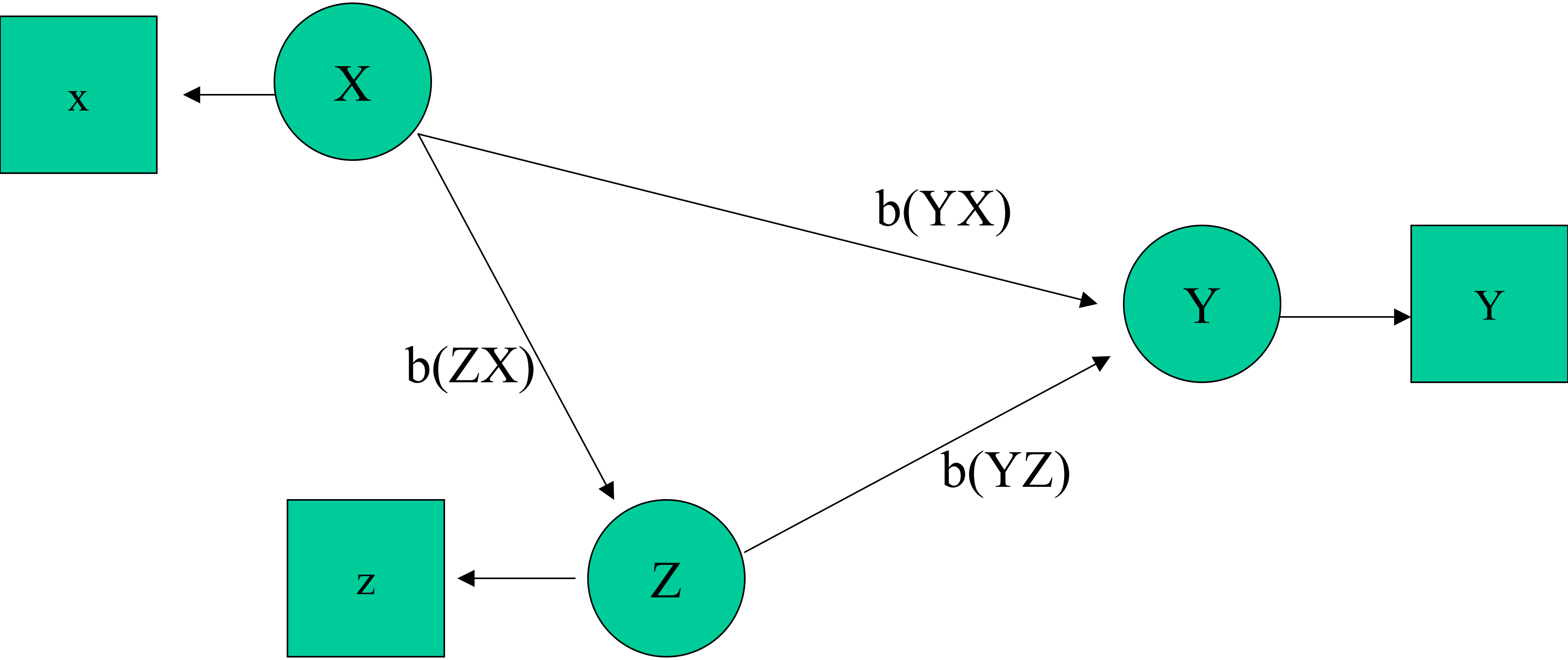
How can we estimate (random) measurement error?

- Repeat the measurement:
  - Y1 = true score + error1
  - Y2 = true score + error2
- Again an elementary causal model with elementary algebra: rY1Y2 = bTY1*bTY2.
- This is not identified by itself, but becomes so if we assume bTY1 = bTY2.
- The measurement b's are also identified when there are three observed indicators, or when there are more variables in the model.

Repeating measurement

- The most elementary design to estimate measurement error is repeating.
- The correlation between two repeated measure is called the test-retest reliability.
- Repeating measures can be at two points in time, but then we must assume that the true score has not changed.
- More commonly the design is aimed at repeating at the same time, under different guises (‘alternate form’).
- This a an extremely common procedure in attitude measurement, but is rarely applied when measuring social structural variables.

Elementary causal model with measurement error



Random measurement and partial effects

- Effects of random measurement error become more complicated when we look at the consequences on direct, indirect and confounding effects.
- Measurement error :
  - In the intervening variable reduces the estimated indirect effect and enlarges the estimated direct effect.
  - In the confounding variable: reduces the estimated spurious association and enlarges the estimated direct effect.
  - In the outcome variable: no change in ratio of direct and indirect effects.
- The first two forms of measurement error lead to **BIAS**, the latter does **NOT**.

Systematic error (invalidity)

- Systematic error arises when measurement error is influenced by other variables, inside or outside the model.
- The various kinds of validity often distinguished in methodological textbooks (construct, content, predictive, discriminant), are not very helpful.
- But the most common textbook definition is: a measure is valid if it measures what you intend to measure (and nothing else).
- The most useful model to think of is the multiple common factor model: indicator is influenced by multiple true scores.

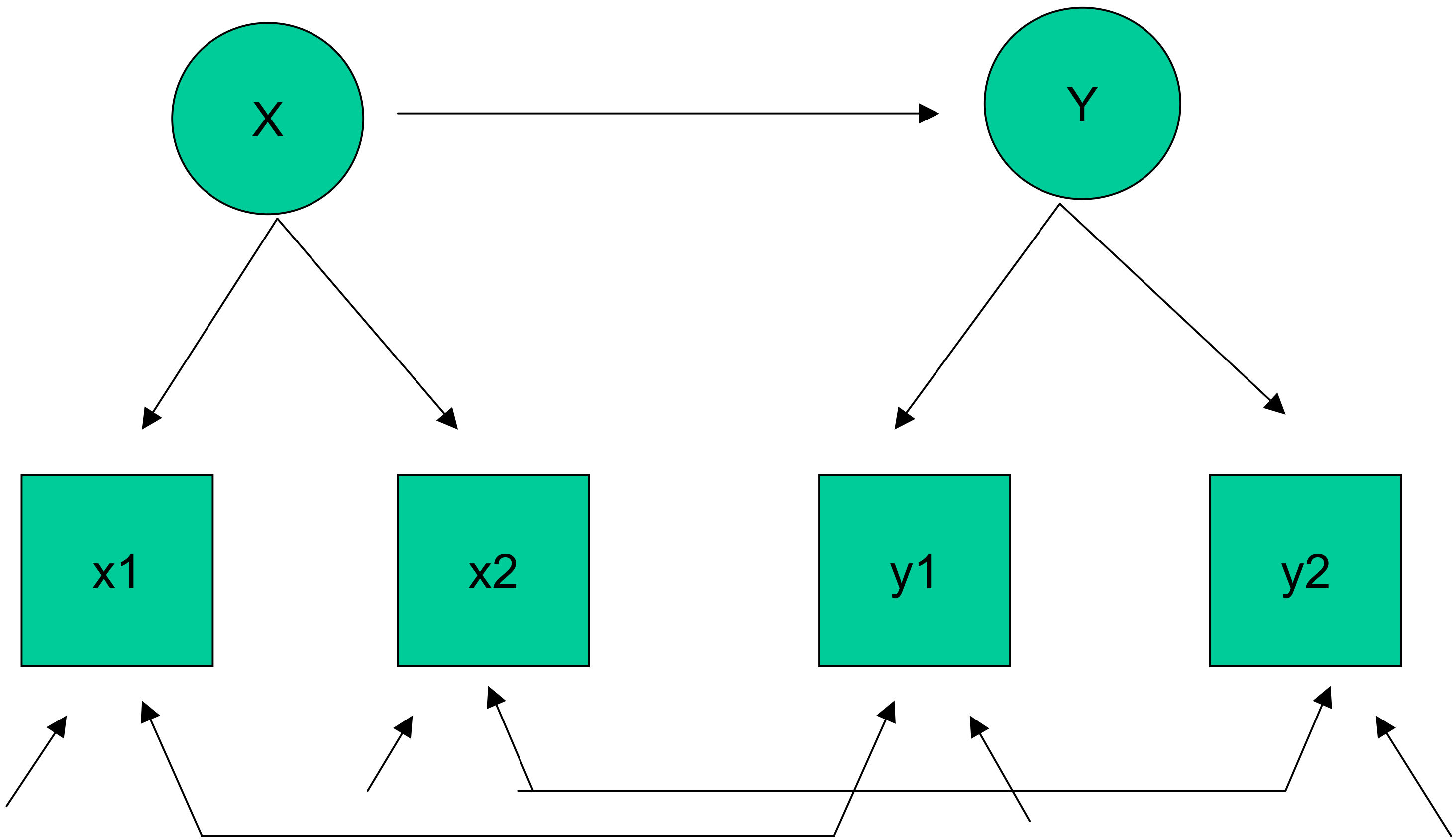
Influences of systematic error

- Influences (bias) of systematic errors cannot easily be summarized in general statements.
- Systematic error increases correlations with some variables, but which ones (inside or outside the model), depends upon the structure of the underlying process and how it is represented in your model.
- Systematic error can in some instances be quite harmless (e.g. constant error).

How can systematic error be detected?

- By repeating the error!
- I.e. repeat the error in measuring another latent variable in the model.
- With repeated systematic error, the size of the error can be estimated using a ‘correlated error term’, or using multiple factor analysis as part of the causal model.
- The is called MTMM (multiple trait, multiple method) modeling.

Elementary MTMM model



An elementary MTMM model – the algebra

- In the most elementary MTMM model we have four indicators for two latent variables: six correlations.
- The coefficients are not identified, unless using restrictive simplifications.
- However, all coefficients become identified, if the MTMM is embedded in a larger causal model.

4. History of stratification research

Generations of research

- B&D is not the beginning of social stratification research; it is commonly regarded as the start of the ***second*** generation.
- Before B&D: first generation
- After B&D: third (and fourth?) generation.

Ancient

- Classical contributions:
  - Marx on social classes
  - Weber on the distinctions between economic classes, status (prestige) groups and political power groups.
  - Durkheim on consensus in representations of society.
- Modern stratification research starts with Sorokin's (1928) "Social and Cultural Mobility". His main conclusion on trends: "trendless fluctuation".

First generation

- Around 1950 sociologists in the ISA planned a series of national stratification surveys, to be held every 10 years:
  - Occupational prestige,
  - Occupational careers,
  - Intergenerational (father-son) occupational mobility.
- When international comparisons appeared (Lipset & Zetterberg 1959; Miller 1960) there were on 3x3 tables (farm, manual, non-manual). Conclusion: intergenerational occupational mobility patterns look “much the same in industrialized societies”.

First generation (2)

- Main tool of analysis were cross-tabulations, that were described using percentages (inflow/outflow).
- First generation analysts avoided association measures (such as the correlation) and studied patterns cell-by-cell.
- Despite prestige being the acclaimed comparative measurement tool, it was very little used by this generation.

Prestige of occupations

- One way to measure status of occupations is by prestige (= popular evaluation). Prestige surveys have been held quite frequently. Main results:
  - Prestige ladders of different countries and different time points look very much alike (the ‘Treiman constant’).
  - There is much consensus about occupations. Essentially, it makes no difference whom you ask. Even student-respondents can produce valid prestige ladders.
- Existing national prestige scales were averaged into the Standard International Occupational Prestige Scale [SIOPS] by Treiman (1977).

Second generation (1)

- Blau & Duncan introduced the following innovations:
  - Detailed occupational measurement.
  - Measurement of occupational status using socio-economic status [SEI].
  - Multivariate causal models using linear (regression and correlation) methods.
- While the status attainment model is multivariate, its expression of association is extremely simple (one number has to tell it all).

Second generation (2)

- The second generation of stratification researchers decided to stage a new round of surveys on the model of the B&D project.
- These surveys were held between 1972 and 1978 in some 15 countries. Famous: OCG-2, Oxford Mobility Enquiry.
- Around 1980 everything was ready for a large scale comparisons of status attainment models.
- It never happened.

What happened?

- Growing discontent with continuous variables as representation of occupational stratification.
- Marx prevailed over Durkheim!
- The rise of log-linear models: a new class of statistical models that can represent association as a multi-parameter phenomenon.

Third generation (1)

- Log-linear (odds-ratio) models were first published in 1975. These models:
  - Require discrete (nominal) data
  - Can separate intrinsic association patterns from developments in marginal distribution (separate social fluidity from structural mobility).
  - Represent association patters as multi-parameter phenomena.
- Comparative research culminates in Erikson & Goldthorpe (1992) “The Constant Flux”.

Third generation (2)

- The favorite measurement instrument for occupational status among third generation researchers was the EGP social class scheme (named after Erikson, Goldthorpe & Portocarero).
- The EGP schema combines occupations, self-employment and supervisory status into 10 (11, 12) classes.