SEM: Simultaneous Equation Models – Introduction and Overview

Harry B.G. Ganzeboom

GSSS PhD Course

October 30 2013

SEM

- SEM can mean:
  - Structural Equation Modelling
  - Simultaneous Equation Modelling
- Brings together:
  - Structural (causal ‘path’) model with multiple dependent (endogenous) variables.
  - Measurement (i.c. common latent factor) model.
- The mathematics of multiple regression analysis applies to both parts.

Advantages

- SEM makes you think about how the world works: causal effects are everywhere.
- SEM makes you aware of the biases that occur through (random and systematic) measurement error and provides tools to diagnose and repair these.
- Also:
  - ML estimation of data with random missings.
  - Constrained estimation: SEM’s can fix coefficient to some specific values, make them equal or restrict them to a certain range.

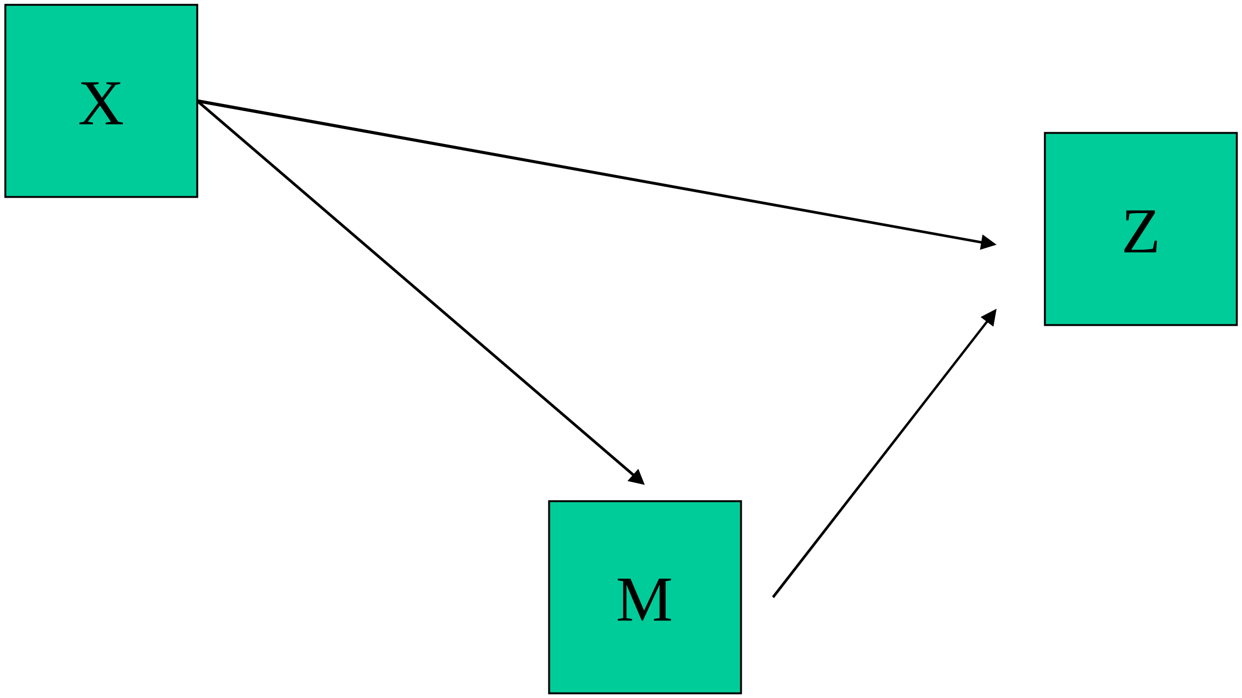
Software

- LISREL (Jöreskog & Sörbom) – SEM models are often referred to as “lisrel-models”, users sometimes as “lisrelites”.
- AMOS
- Mplus
- Stata12
- I use LISREL and Stata12, but my impression is that Mplus is a bit more powerful.

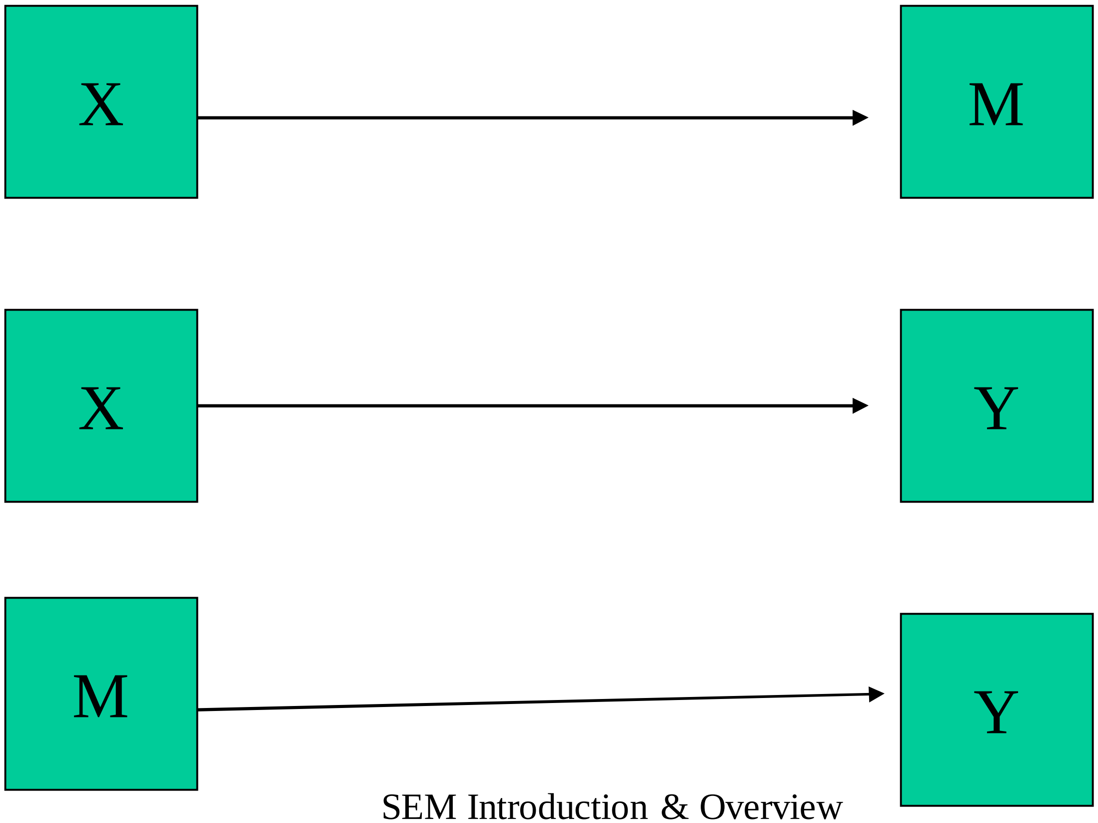
The world as a correlation matrix

- Although SEM’s are not restricted to covariances and correlations, the classical models and applications are.
- In an SEM state of mind, one sees the world summarized in a correlation matrix; the researcher’s task is to invent a set of (simple) mathematical equations that will reproduce this matrix.
- I find it more convenient to think in terms of correlations (standardized) in stead of covariances.
- The algebra of correlations and causal effects is fairly simple. You do not need regression analysis to estimate path coefficients, the path-analytic decomposition will do!

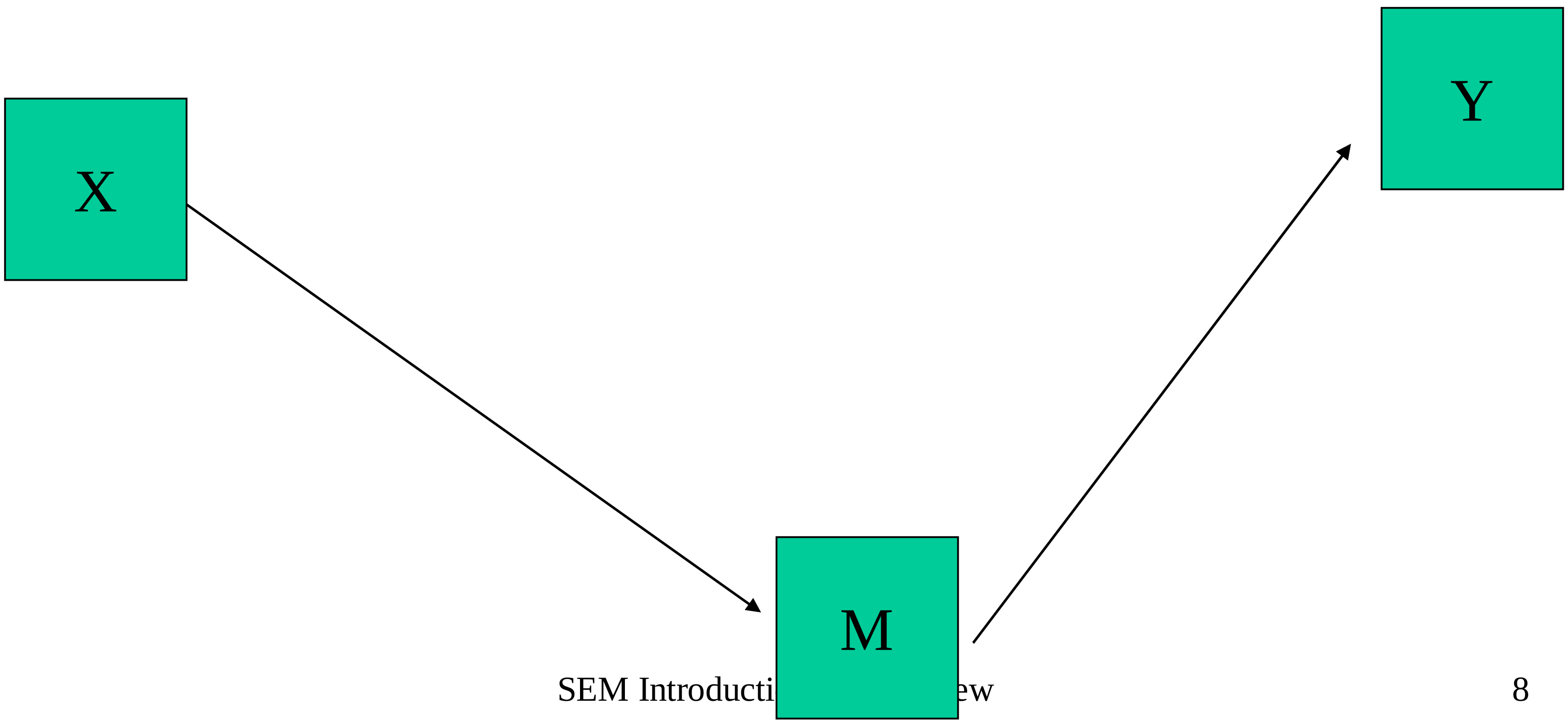
The elementary causal model



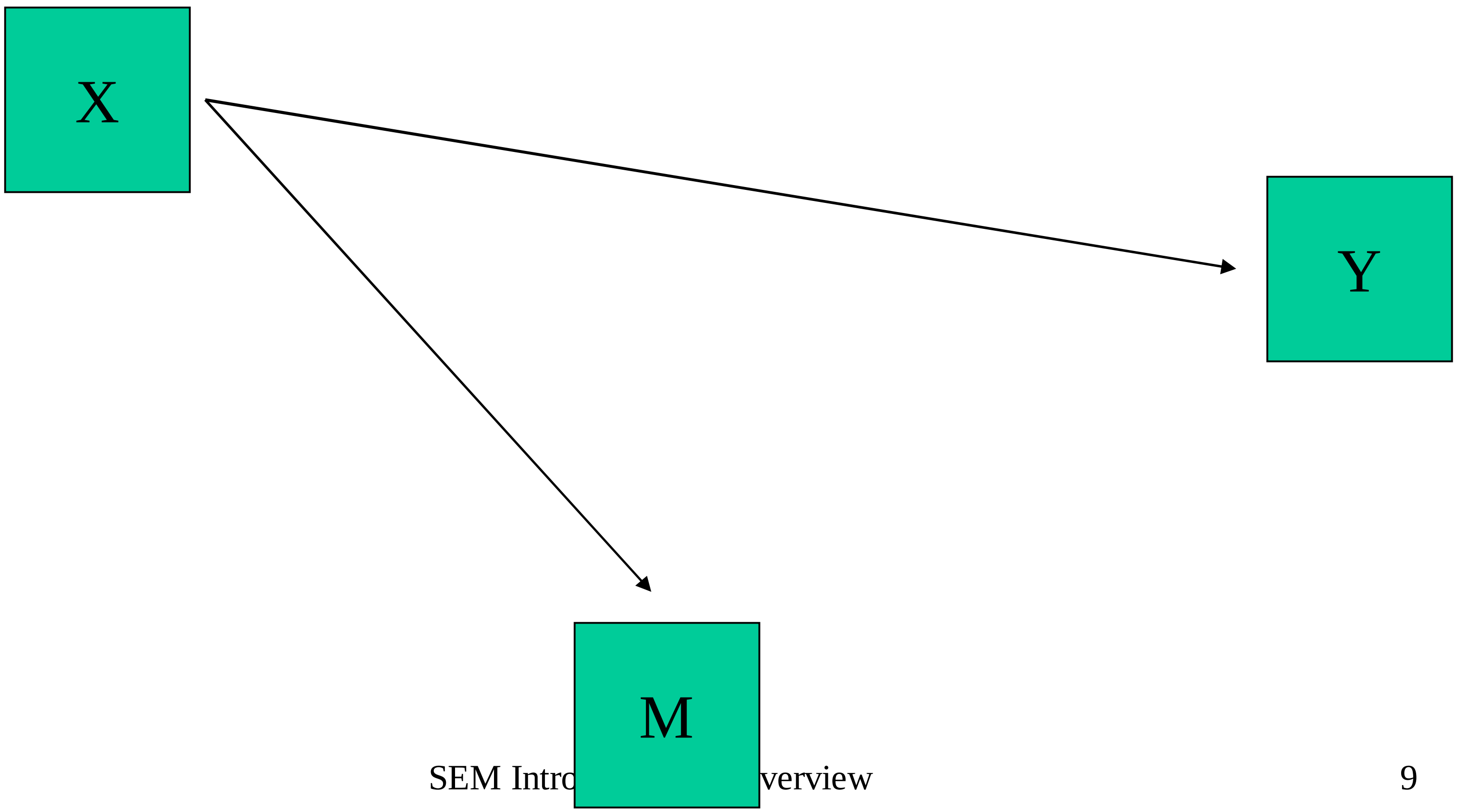
Direct effects



Indirect effect



Confounding effect



The path-analytic theorem

- Total correlation =
  - ***Direct effect + indirect effects + confounding effects.***
- Indirect effects are the multiplication of the two direct effects (in a chain).
- Confounding effects are the multiplication of the two direct effects (in a fork).
- Notice that while the definition and calculation of confounding and indirect effects is fairly similar, their causal interpretation is radically different:
  - Indirect effects inform you how (via which mechanism) X causes Y;
  - Confounding effects inform to what extent the correlation between X and Y is NOT causal (but spurious).

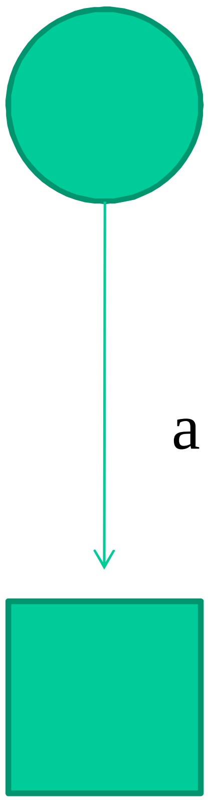
Identification

- In the elementary causal model there are three correlations (known quantities) and three unknown quantities (coefficients).
- This generates a system of three algebraic equations that can simply be solved (by hand).
- The solution is identical to what we would find using (multiple) regression (2 (!) equations).
- In non-recursive models it is generally true that we have as many unknowns as equations, and with some work, can solve for the unknowns.

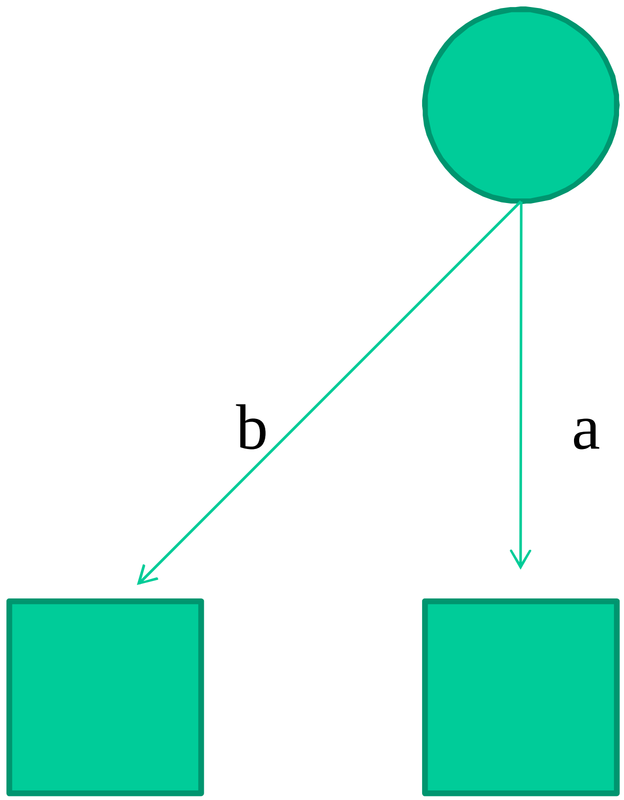
Measurement models

- Measurement can be interpreted as a causal model in which a latent variable causes the response on an observed variable.
- → We see the reality in our observed variables always with some measurement error (Plato).
- We can estimate the measurement error when we repeat the measurement and compare the indicators.
- If we do not have repeat measures, we cannot know the amount of measurement error, but it is still there.
- Measurement error in a model with two measures is not identified as such (but see below), but a latent variable model with three indicators is exactly identified, much in the same way as we can solve for the coefficients in the elementary causal model (three correlations and three unknown coefficients).

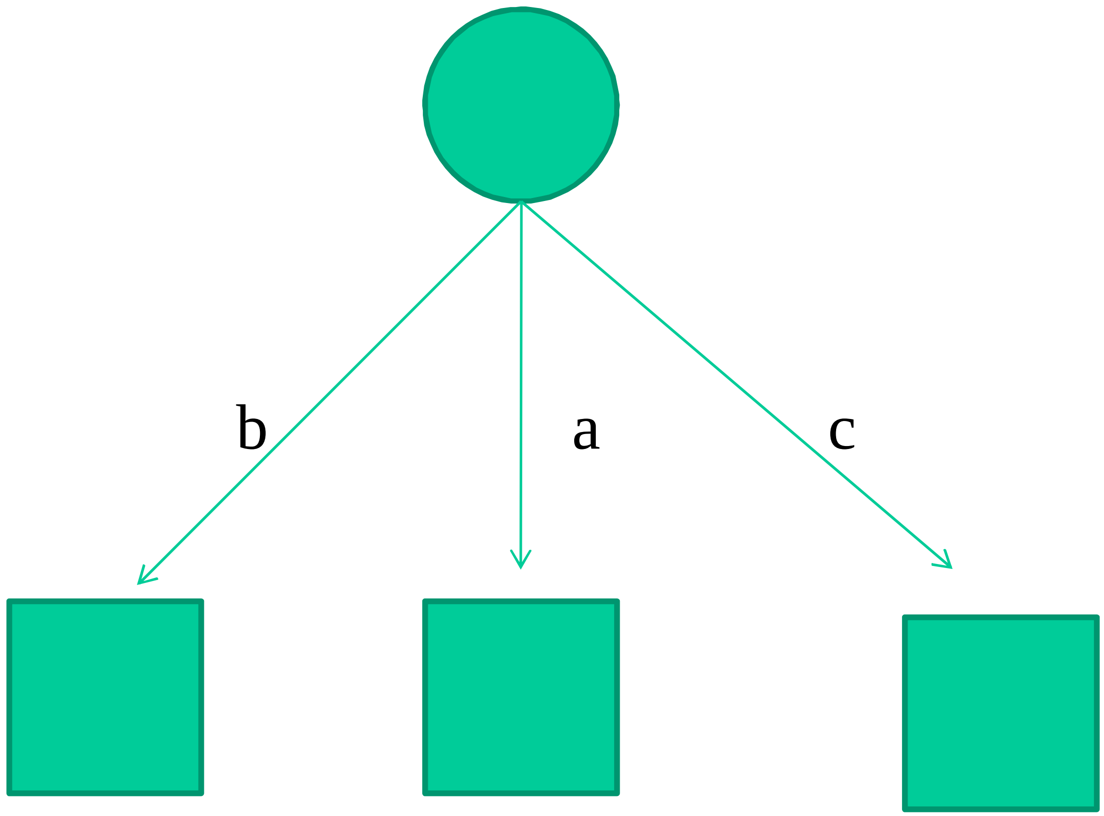
Elementary measurement model



Elementary measurement model



Elementary measurement model



Elementary measurement model

- The measurement model with three indicators is exactly identified: we can find out the quality of each indicator separately (with respect to random measurement error.)
- However, when estimated in a larger model with auxiliary variables – two indicators will often do.

Putting it together

- The elementary causal model and the measurement model are both SEM's, but the real SEM arises when we combine them in a single model.
- Note that if we combine measurement model and causal model, it is (mostly) not necessary anymore to have three indicators for each latent variable: two is enough for identification.

Random measurement error

- Random measurement error (or: Unreliability) arises as if by a random process: it is unpredictable when and how much deviation from the true score will arise for each individual.
- Random error makes measures unreliable (or: unstable): it leads to different answers all of the time.
- With SEM common factor model we can estimate *how much* error occurs, but also diagnose *where* it occurs.

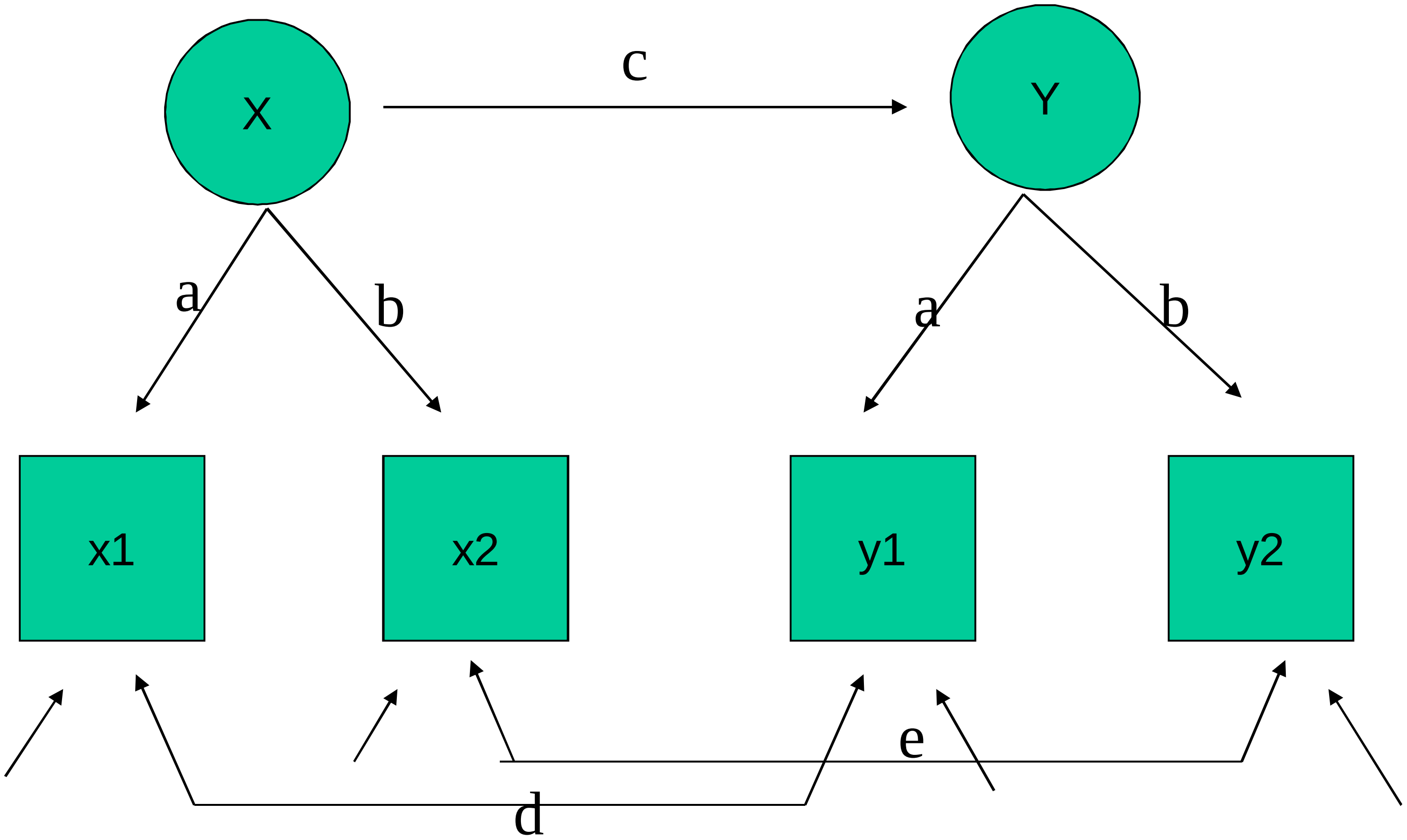
Systematic measurement error

- Some kind of measurement error arise systematically, the deviation from the true score has some consistency (within persons, between measures).
- Systematic measurement error is also known as invalidity or bias.
- We can trace systematic errors by repeating the error:
  - Random error: repeat the measurement
  - Systematic error: repeat the error.

Correlated error

- If we have two measures that have the same (=systematic) error, this arises as correlation between the measures (even if the two measures do not have a true score in common).
- Systematic measurement modeling is just a variety of (multiple) common factor analysis.
- MTMM models: Multiple Traits, Multiple Methods – is a traditional name for separating random error from systematic (‘method’) error.

SEM / MTMM model



SEM and factor analysis

- Another perspective on SEM is that it adds a causal structure to the correlations among the common factors.
- As a matter of fact, it is useful to run any SEM models with multiple indicators as such a common factor model, either exploratory (SPSS) or confirmatory.
- This model informs about the measurement part, without being confounded by problems in the causal part.

How to do it -- outline

- Generate the correlation / covariance matrix to analyze. You can also analyze individual data.
- Write up your causal / measurement model as a diagram.
- Write up your causal / measurement model as a set of linear equations (one for each dependent variable).
- Program the equations in a SEM program.
- Assess how the reproduced correlations fit the observed correlations.
- Adjust ('modify') the model by allowing more effects or trim superfluous ones.