METHODS OF QUANTITATIVE DATA ANALYIS
MSR Course, 2011-2012

Harry B.G. Ganzeboom

Lecture 4c: SEM models

May 4 2012

SEM

- SEM can mean:
  - Structural Equation Modelling
  - Simultaneous Equation Modelling
- Brings together:
  - Multiple (multiple) regression models in a single causal (structural model)
  - Regression models for causal effect with measurement models (I.c. comon factor model).

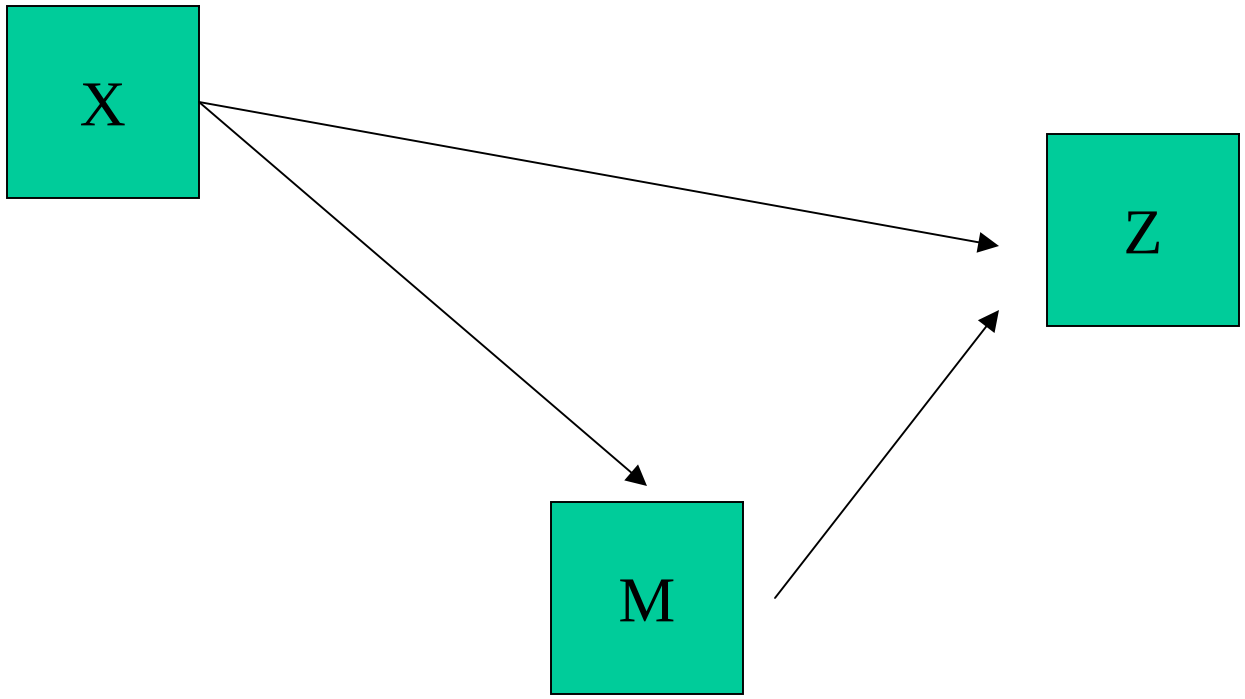
Advantages

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

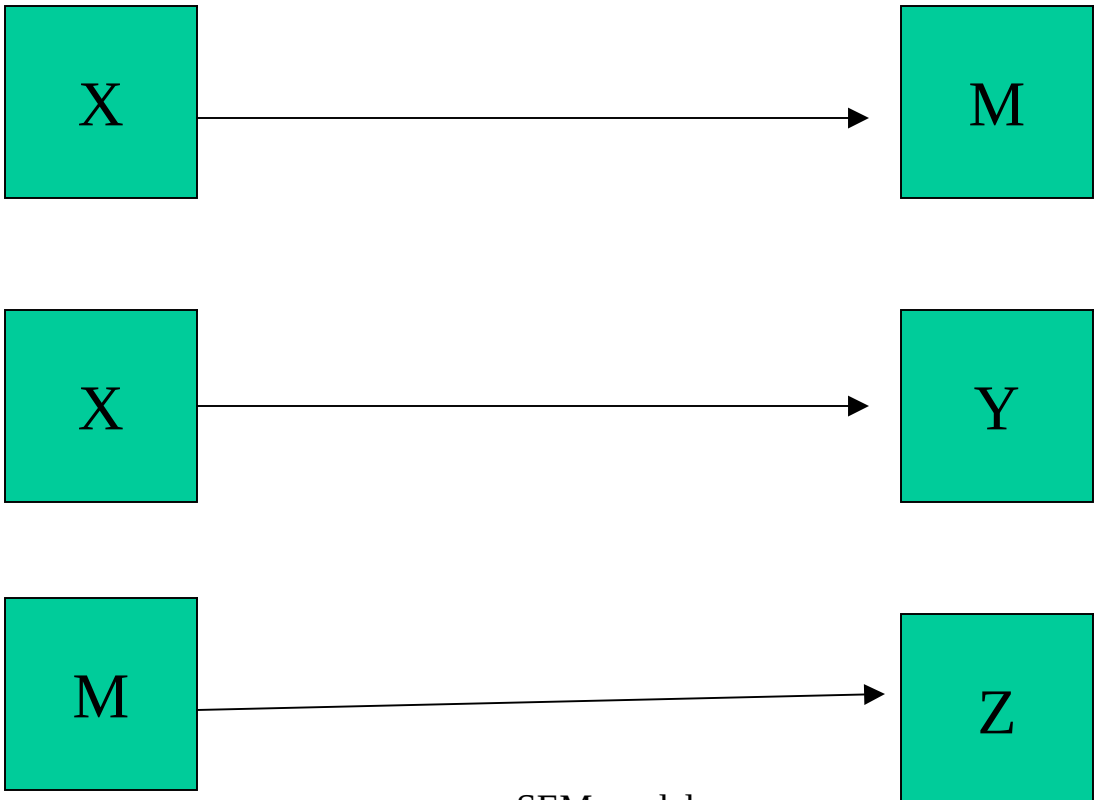
The world as a correlation matrix

- 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 research task is to invent a set of (simple) mathematical equations that will reproduce the matrix.

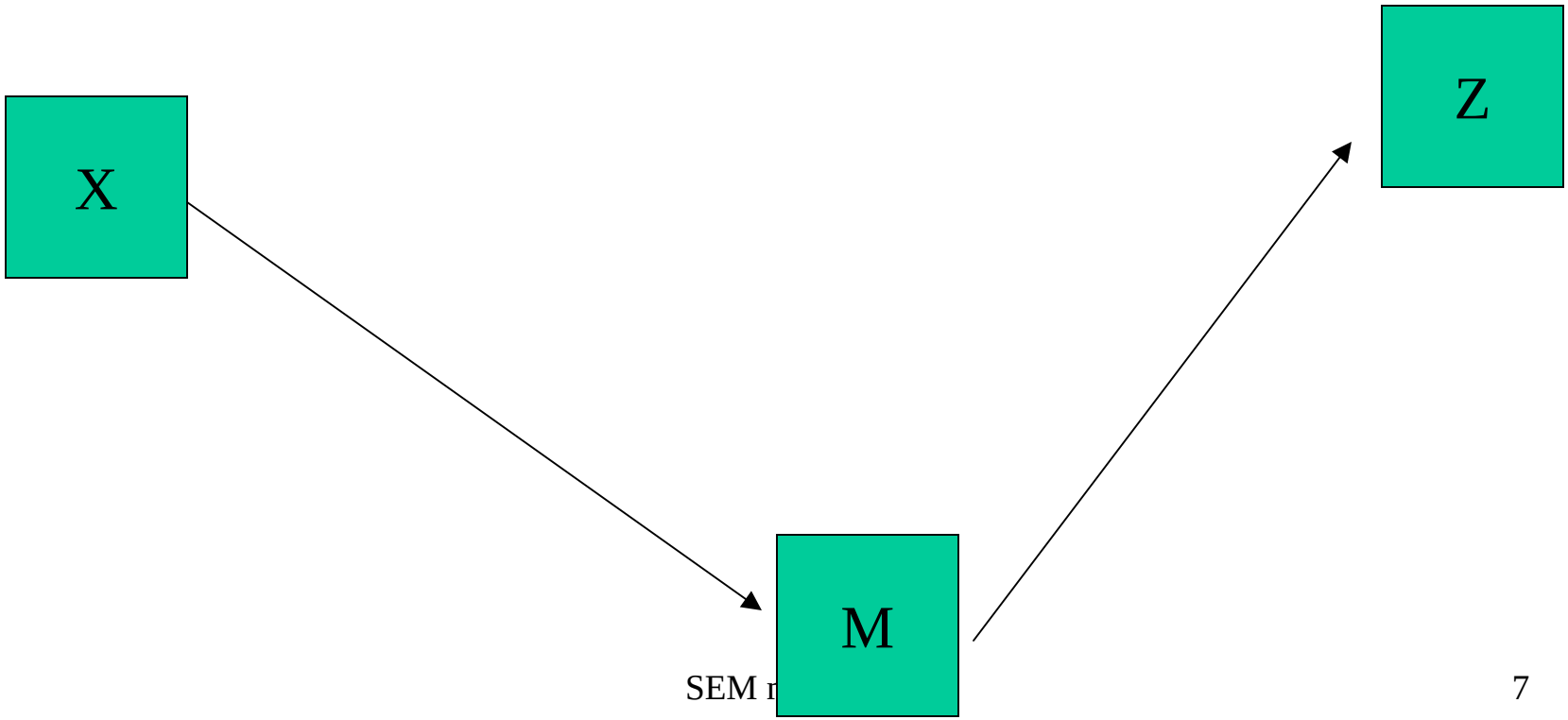
The elementary causal model



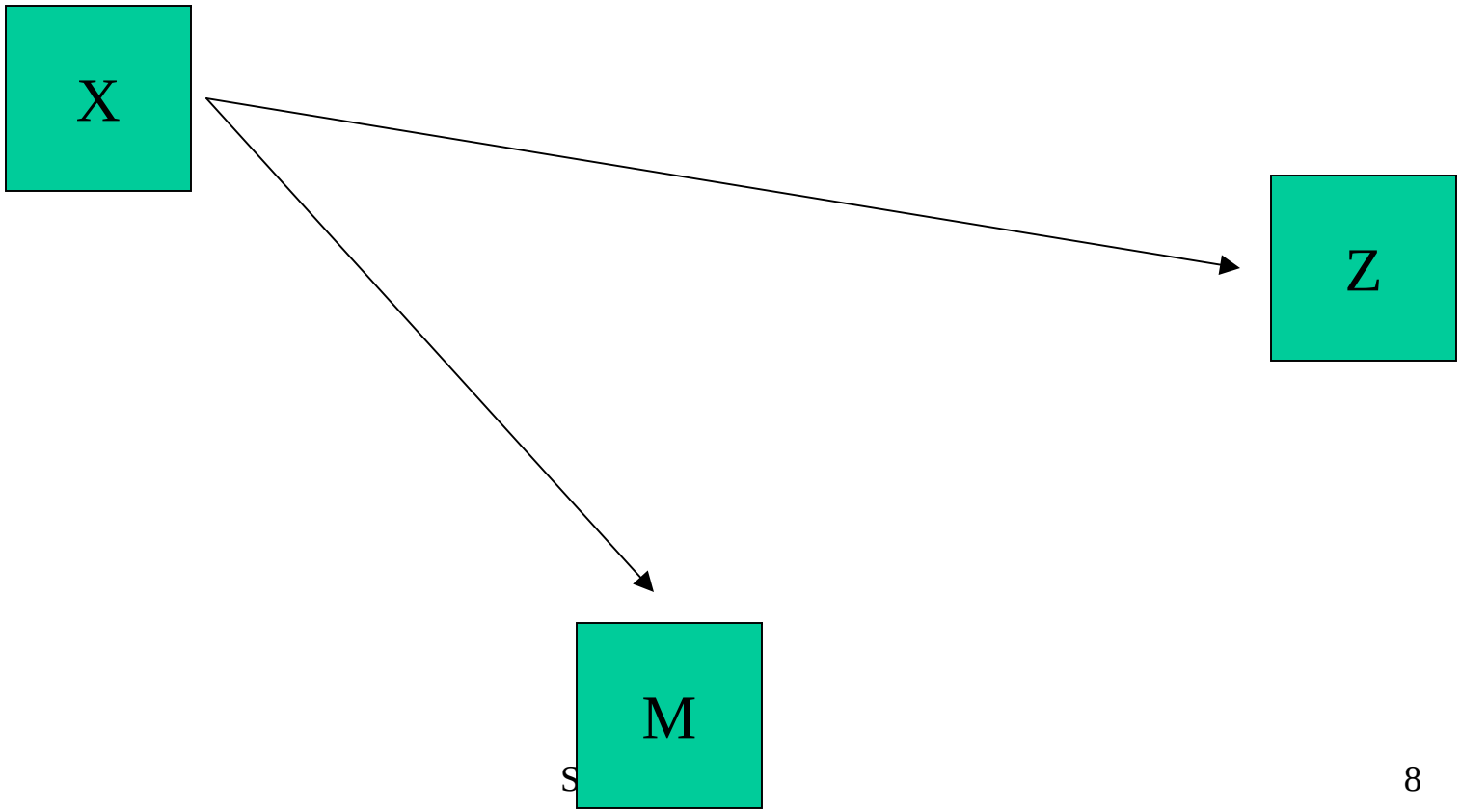
Direct effects



Indirect effect



Confounding effect



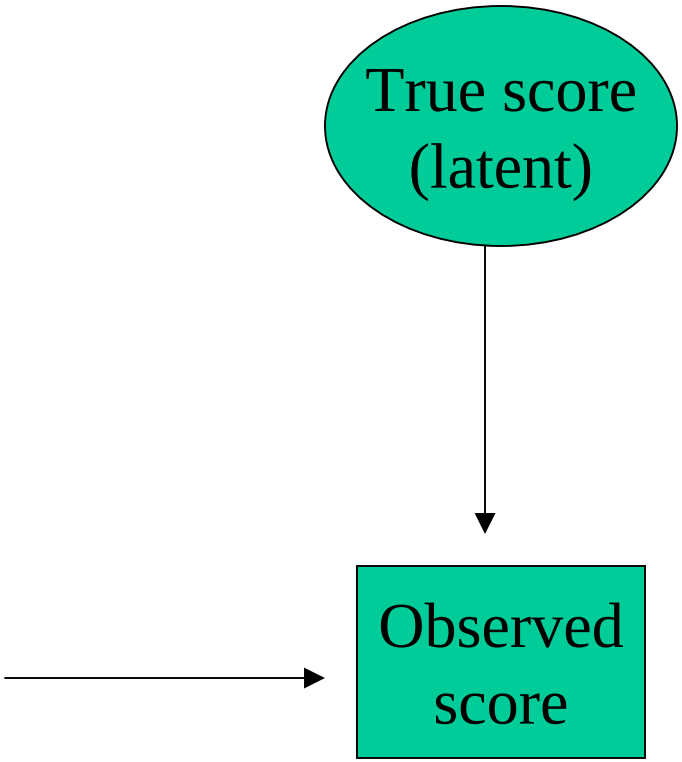
The path-analytic theorem

- Total correlation =
  - ***Direct effect + indirect effects + confounding effects.***
- Indirect effects are the multiplication of the two direct effects.
- Confounding effects are the multiplication of the two direct effects.
- 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 influences Y;
  - Confounding effects inform to what extent the correlation between X and Y is NOT causal (but spurious).

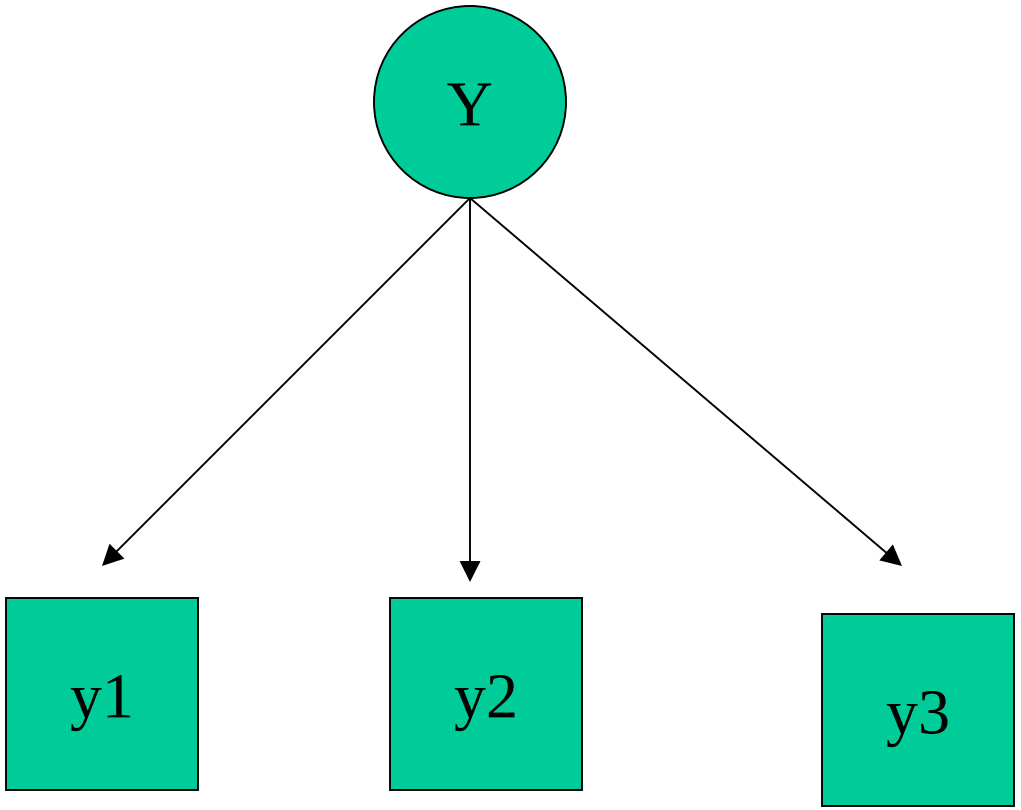
Identification

- In the elementary causal model there are three correlation (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.
- 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



The elementary measurement model



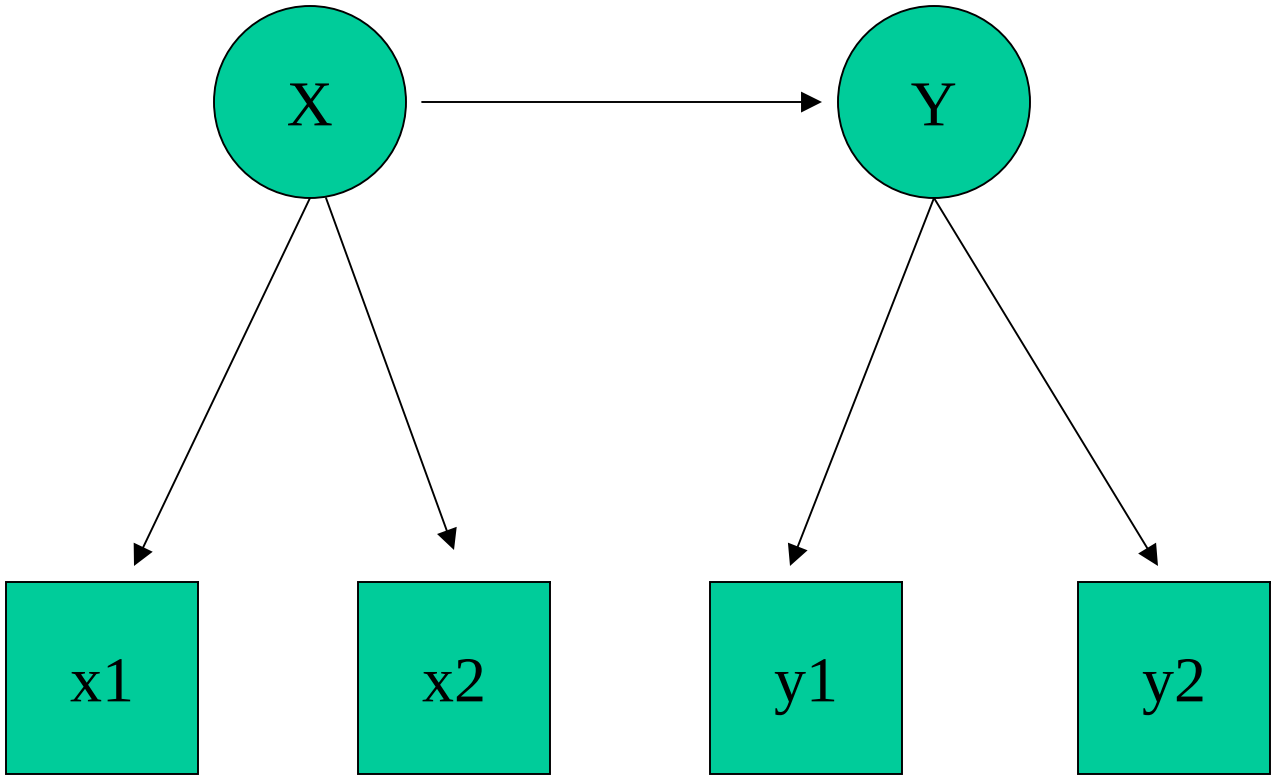
Measurement models

- 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 with some measurement error.
- We can estimate the measurement error when we repeat the measurement and generate independent measurement error.
- 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 model with three indicators is exactly identified, much in the same way as we can solve for the coefficients in the elementary causal model.

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.

A two factor model



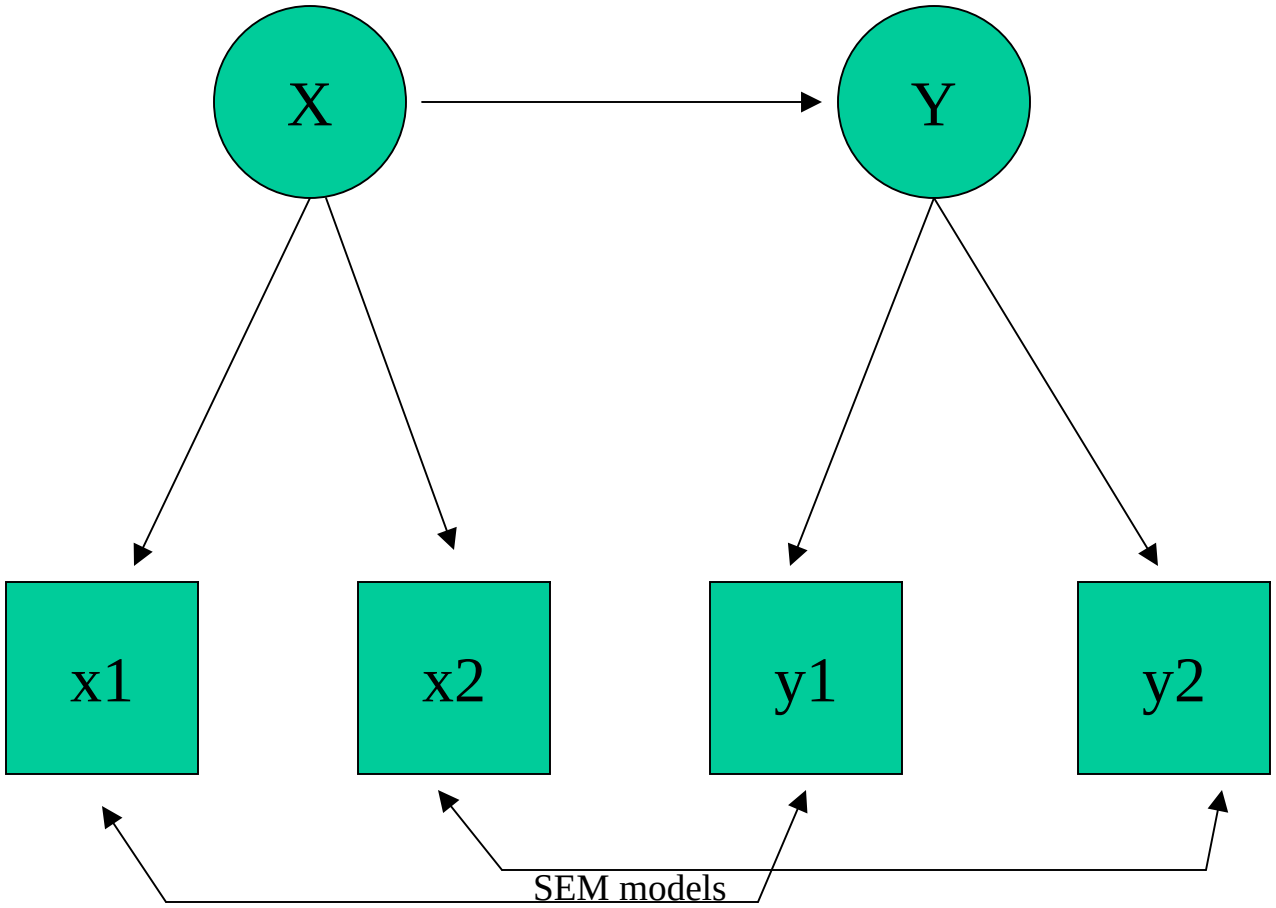
Identification in the two factor model

- In the two-factor model with two indicators per construct we have 5 unknowns and 6 correlation: the model is identified.
- So: by combining causal and measurement analysis, we can reduce the number of necessary indicators.
- And still identify the amount of random measurement error.

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 not find out *when* it occurs.

A two factor model with correlated error



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 measure 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 modelling 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.

Software

- LISREL (Jöreskog & Sörbom) – SEM models are often referred to as “lisrel-models”, users as “lisraelites”.
- Mplus
- AMOS
- Stata12