# SELECTED QUANTITATIVE METHODS
# MSR-course 2012-2013

Lecture 1a: Retrospect & Prospect

September 4 2012

# Some rules

- We meet on Tuesday / Friday
- Hours?
- There is a weekly assignment: to be submitted on Monday / Thursday evening.
- Final assignment (to be graded): Redo 2 assignments.
- Every week you also have to submit an ***interesting*** question. These are also graded !! I will (try to) answer these in writing.
- Depending upon interest and (my) capacity, assignments can be individualized.
- You can work in SPSS and/or Stata.

# Evaluations of previous courses

- SQM: No criticisms!!
- MDA: not enough feedback (on assignments) / too much study load.
- What can I / we do about this?

# SMC (=SQM)

- Course needs to be on advanced masters level.
- It can be adaptive: students state (collective) preferences, teaching can be individualized (tutoring, assignments).
- My proposed options are:
  - SEM (see course on Linear Structural Models)
  - Discrete choice models (see last year SQM).

# SEM

- Simultaneous Equation Models:
  - Causal model: A system of regression equations (multiple dependent variables)
  - Measurement model: Causal model is integrated with latent variable (factor) analysis.
- Also:
  - Constrained estimation: zero, equality and scaled constraints on coefficients.
  - FIML modelling of missing values (the best way to do pairwise analysis).
  - Extensions to panel and time series models.
- Can be done in Stata or stand-alone programs: LISREL, Mplus, Amos.

# The limitations

- SEM models assume that the world can be summarized in variables that are expressed in a (metric) score with a mean and a standard deviation, that are related via correlations.
- For statistical inference it must also often be assumed that the world is a multivariate ***normal*** distribution.

# What is NOT covered by SEM?

- Discrete variables are problematic in SEM analysis – in particular as dependent variables – although there are adaptations available.
- Interactions can also be problematic, in particular with a continuous modification variable.

# Discrete choice models

- Discrete choice = choose 1 of 2 (binomial) or 1 from many (multinomial).
- Binomial and Multinomial Regression (SPSS: Logistic and NomReg).
- Two more advanced / complicated situations:
  - When the alternative are rank-ordered (in a single dimension): ordered multinomial regression (SPSS: PLUM).
  - When the alternatives are scaled (in a single or multiple regression): conditional multinomial regression.
- While ordered and scaled multinomial regression models are more complicated to handle than ordinary multinomial regression, the interpretation is less complicated (and the statistical power higher).

# Conditional (scaled) multinomial regression

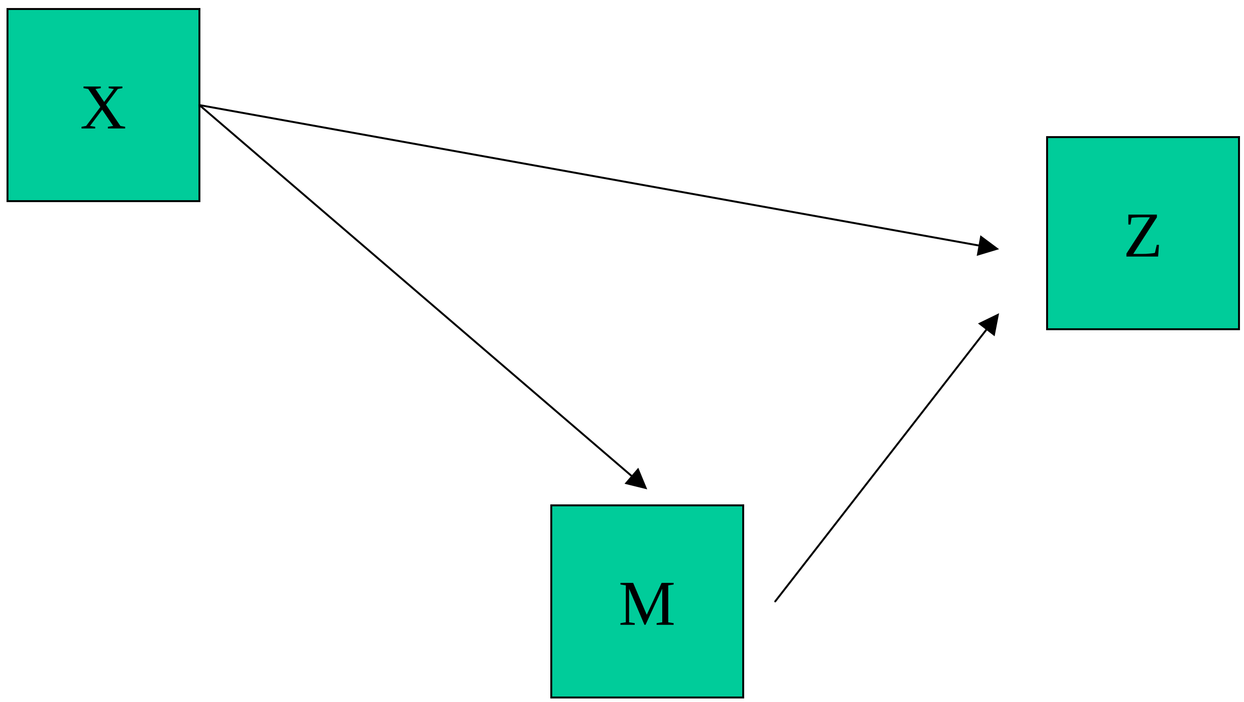
- Are estimated on a person-choice data-matrix (very similar to a person-time data-matrix).
- Independent variable apply both to characteristics of persons (choosers) and the alternatives (choices).
- This makes the model very applicable in all sorts of important situations.

# SEM

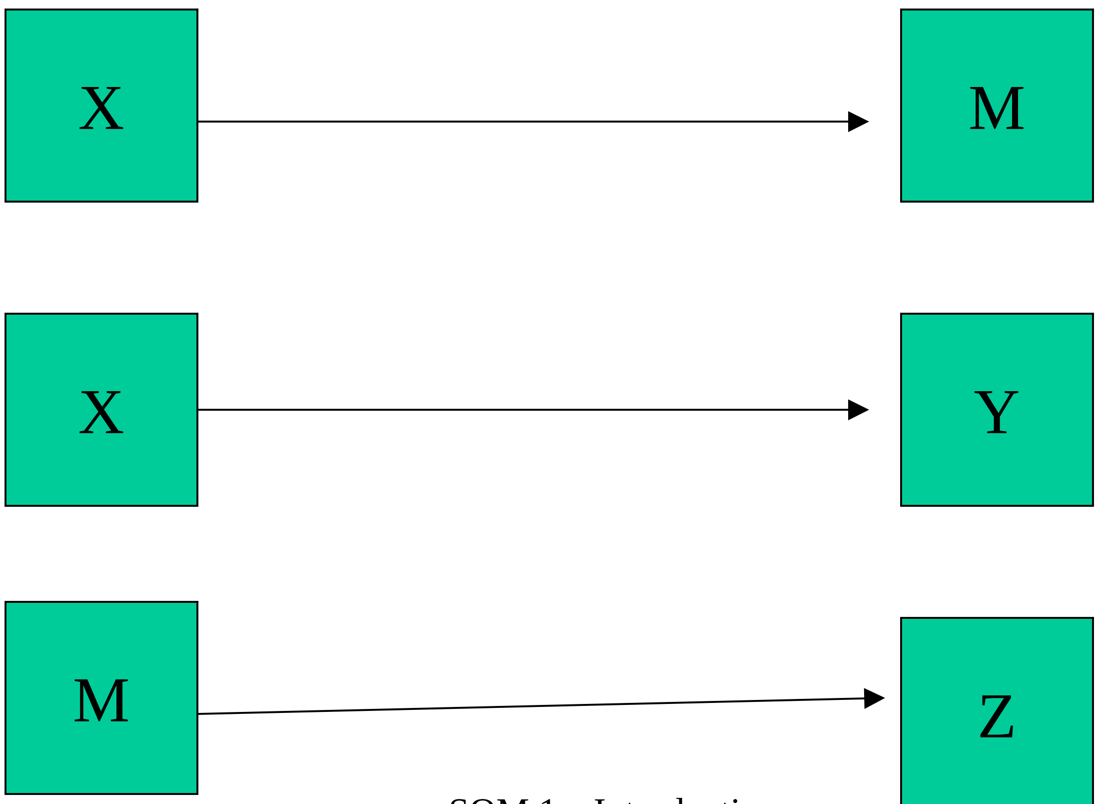
# The world as a correlation matrix

- Although SEM's are not restricted to covariances and correlations, all 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.

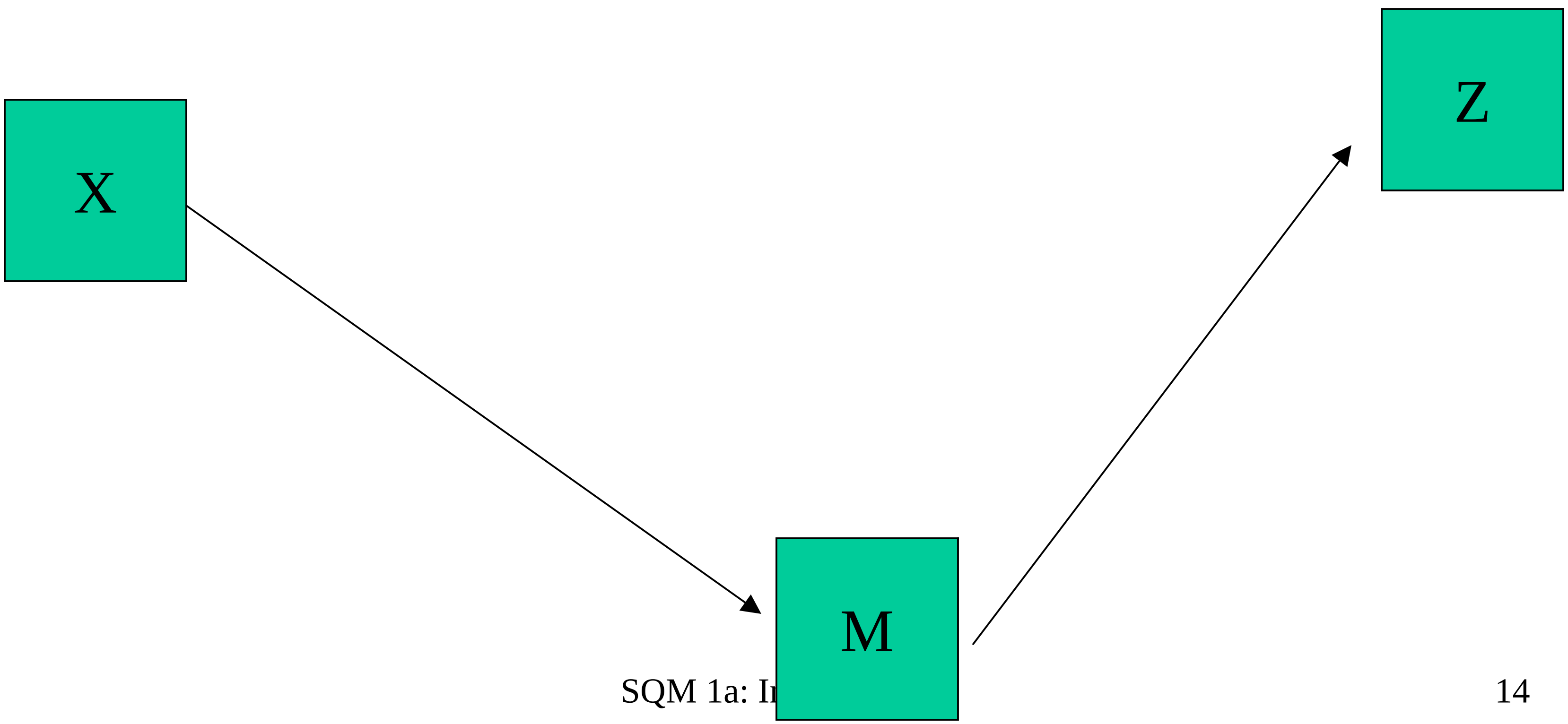
# The elementary causal model



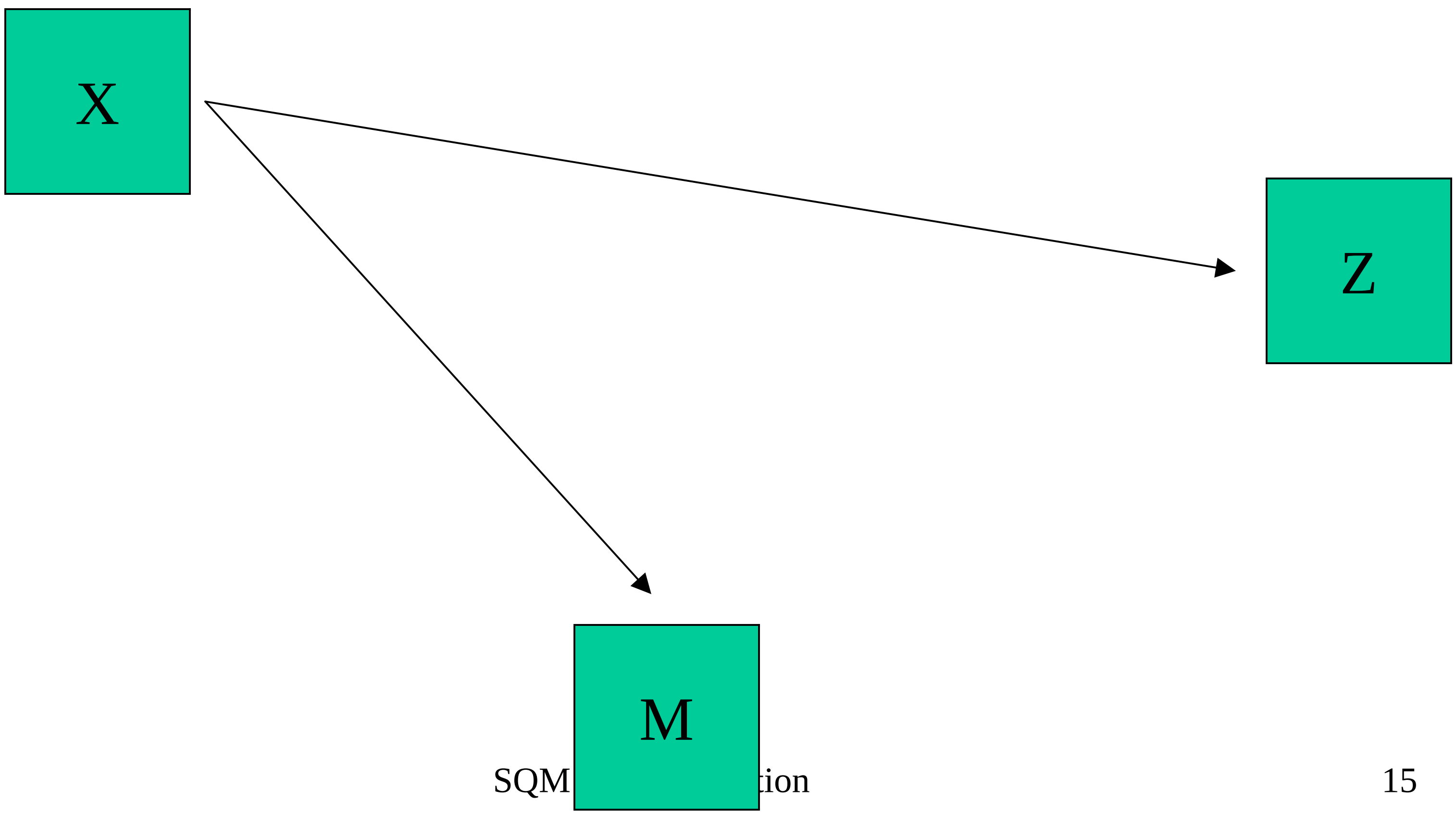
# Direct effects



# Indirect effect



# Confounding effect



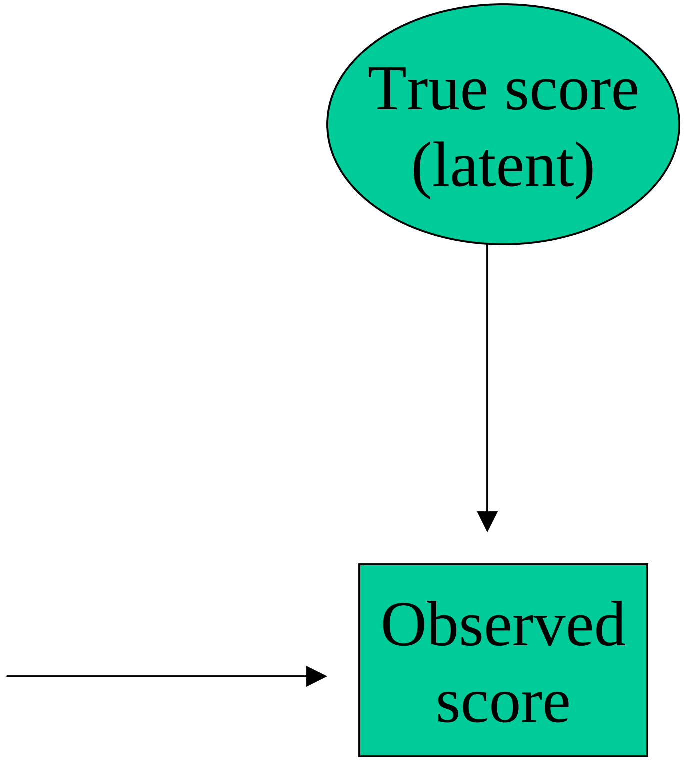
# The path-analytic theorem

- 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).

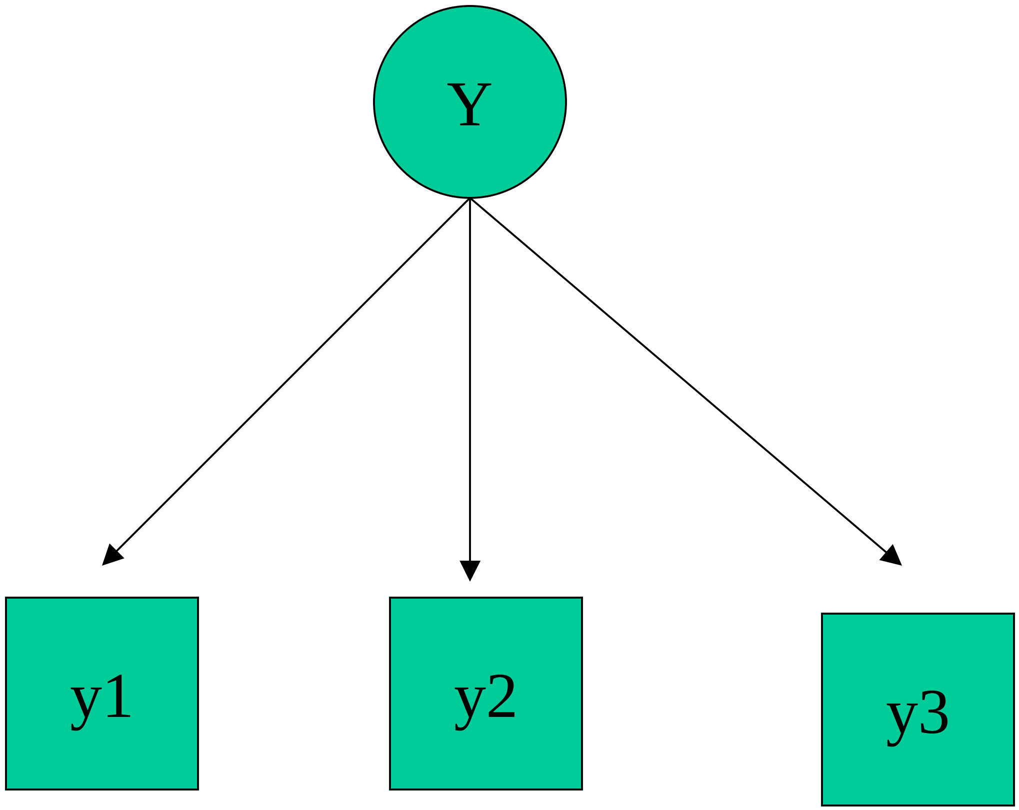
# Identification

- In the elementary causal model there are three 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.
- 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



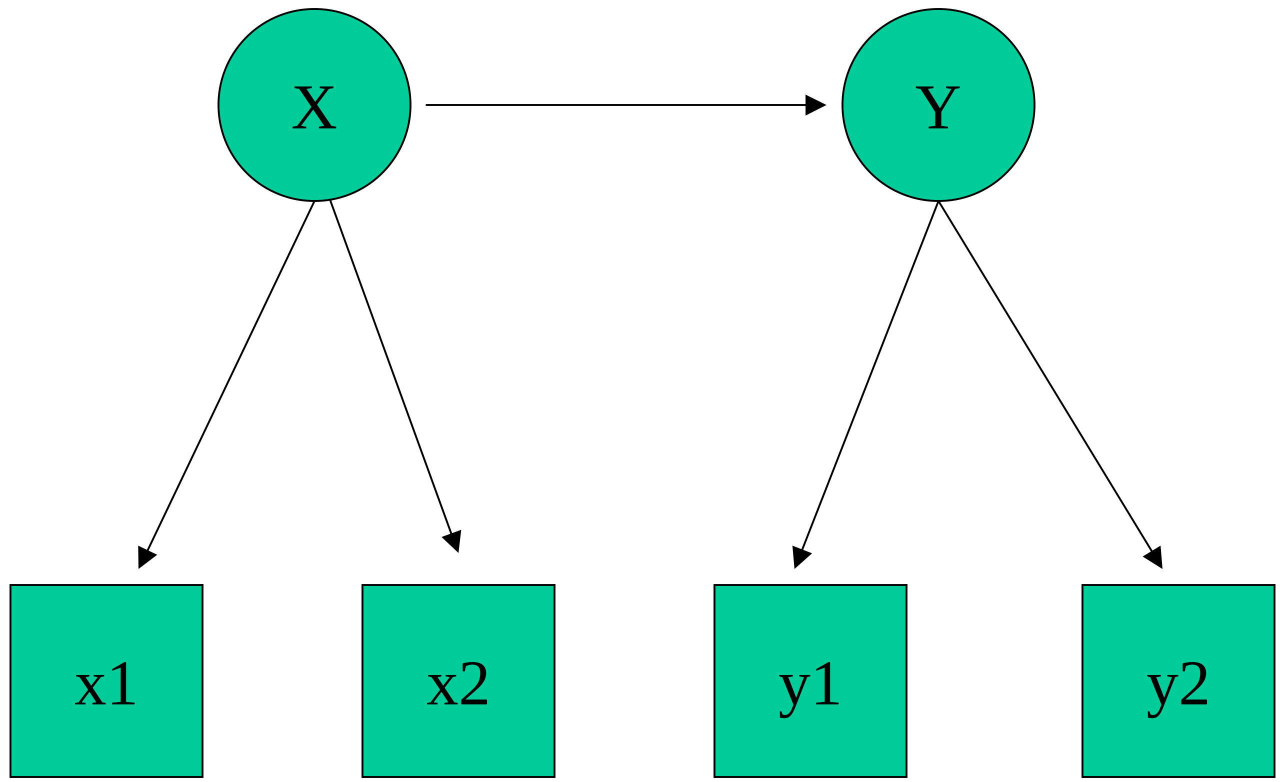
# Measurement models

- 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 three 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



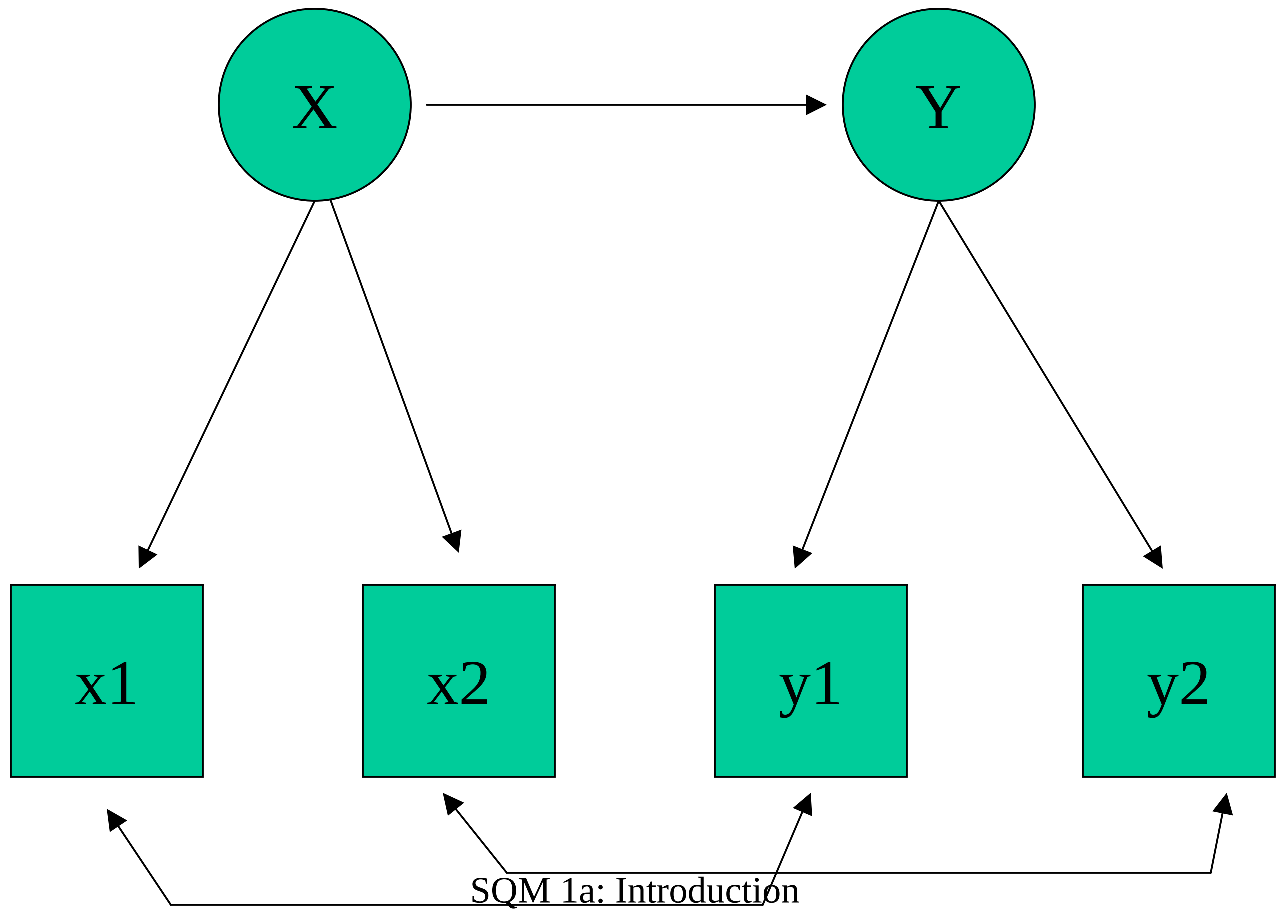
# Identification in the two factor model

- In the two-factor model with two indicators per construct we have 5 unknowns and 6 correlation: the model is (over)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 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 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.