NCRM National Centre for Research Methods

Structural Equation Modeling: what is it and what can we use it for?

Professor Patrick Sturgis

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
UNIVERSITY OF Southampton
MANCHESTER 1824 The University of Manchester
THE UNIVERSITY OF EDINBURGH

What is SEM?

What is SEM?

- SEM is not one statistical ‘technique’

What is SEM?

- SEM is not one statistical ‘technique’
- It integrates a number of different multivariate techniques into one model fitting framework

What is SEM?

- SEM is not one statistical ‘technique’
- It integrates a number of different multivariate techniques into one model fitting framework
- It is an integration of:
  - Measurement theory
  - Factor (latent variable) analysis
  - Path analysis
  - Regression
  - Simultaneous equations

Useful for Research Questions that..

Useful for Research Questions that..

- Involve complex, multi-faceted constructs that are measured with error

Useful for Research Questions that..

- Involve complex, multi-faceted constructs that are measured with error
- That specify ‘systems’ of relationships rather than a dependent variable and a set of predictors

Useful for Research Questions that..

- Involve complex, multi-faceted constructs that are measured with error
- That specify ‘systems’ of relationships rather than a dependent variable and a set of predictors
- Focus on indirect (mediated) as well as direct effects of variables on other variables

Also Known as

Also Known as

- Covariance Structure Analysis

Also Known as

- Covariance Structure Analysis
- Analysis of Moment Structures

Also Known as

- Covariance Structure Analysis
- Analysis of Moment Structures
- Analysis of Linear Structural Relationships (LISREL)

Also Known as

- Covariance Structure Analysis
- Analysis of Moment Structures
- Analysis of Linear Structural Relationships (LISREL)
- Causal Modeling

Software for SEM

- There are a lot of software packages that can fit SEMs

Software for SEM

- There are a lot of software packages that can fit SEMs
- The original and best known is Lisrel, developed by Joreskog and Sorbom

Software for SEM

- There are a lot of software packages that can fit SEMs
- The original and best known is Lisrel, developed by Joreskog and Sorbom
- Mplus, EQS, Amos, Calis, Mx, SEPATH, Tetrad, R, stata

Software for SEM

- There are a lot of software packages that can fit SEMs
- The original and best known is Lisrel, developed by Joreskog and Sorbom
- Mplus, EQS, Amos, Calis, Mx, SEPATH, Tetrad, R, stata
- Some have downloadable student versions

SEM can be thought of as

Path Analysis

using

Latent Variables

What are Latent Variables?

What are Latent Variables?

- Most social scientific concepts are not directly observable, e.g. intelligence, social capital

What are Latent Variables?

- Most social scientific concepts are not directly observable, e.g. intelligence, social capital
- This makes them hypothetical or ‘latent’ constructs

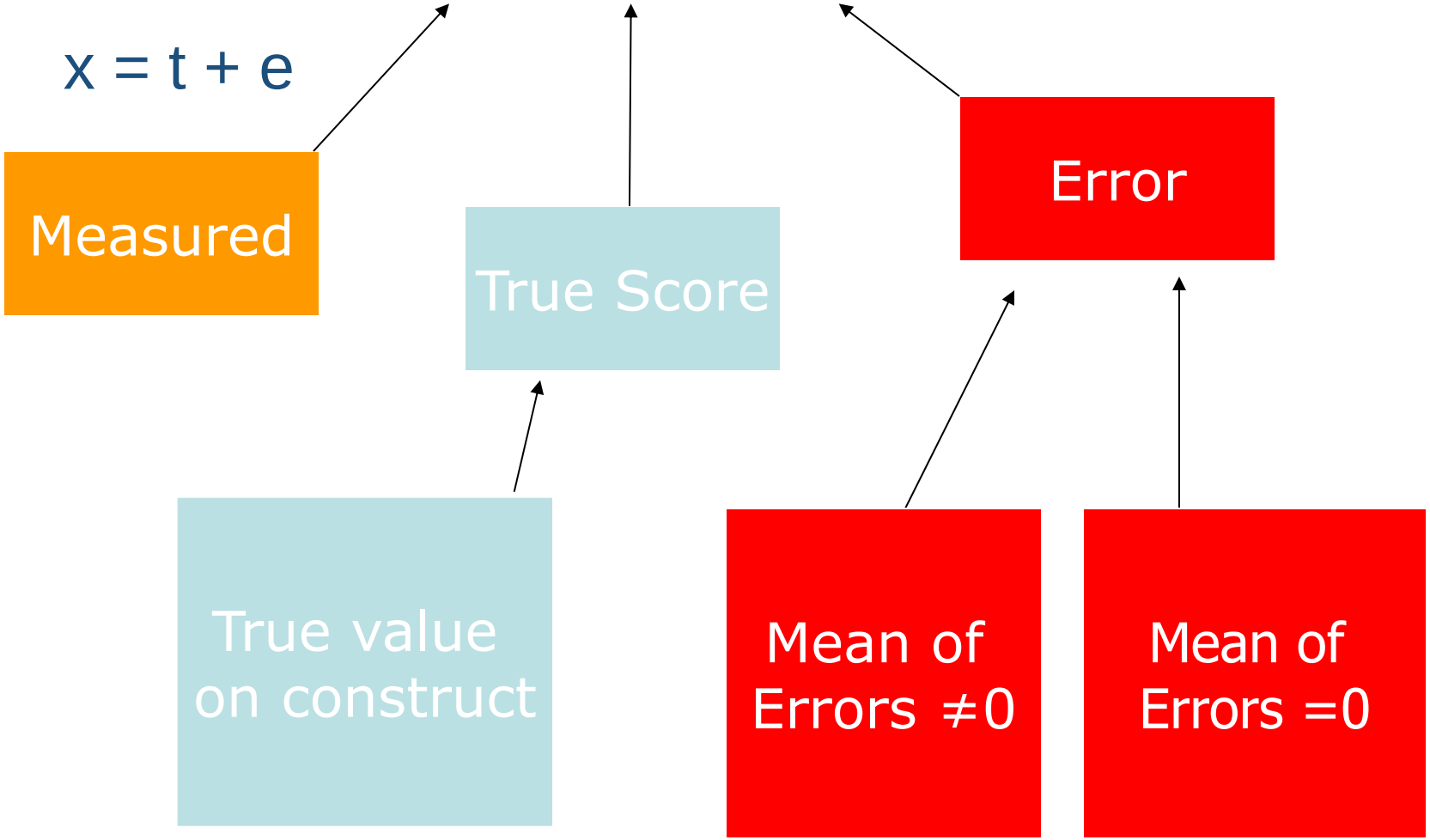
What are Latent Variables?

- Most social scientific concepts are not directly observable, e.g. intelligence, social capital
- This makes them hypothetical or ‘latent’ constructs
- We can measure latent variables using observable indicators

What are Latent Variables?

- Most social scientific concepts are not directly observable, e.g. intelligence, social capital
- This makes them hypothetical or ‘latent’ constructs
- We can measure latent variables using observable indicators
- We can think of the variance of a questionnaire item as being caused by:
  - The latent construct we want to measure
  - Other factors (error/unique variance)

True score and measurement error

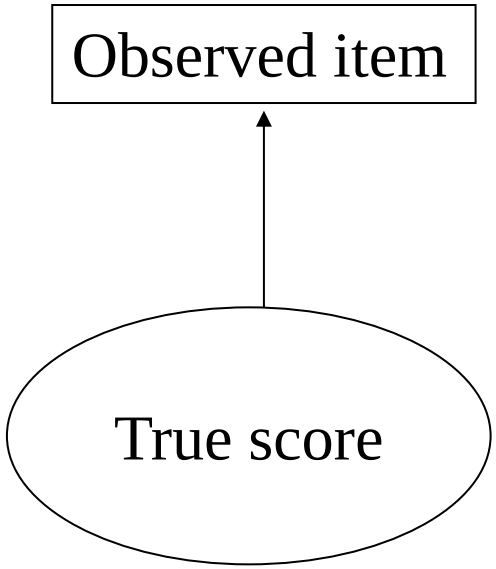


$$X = t + e$$

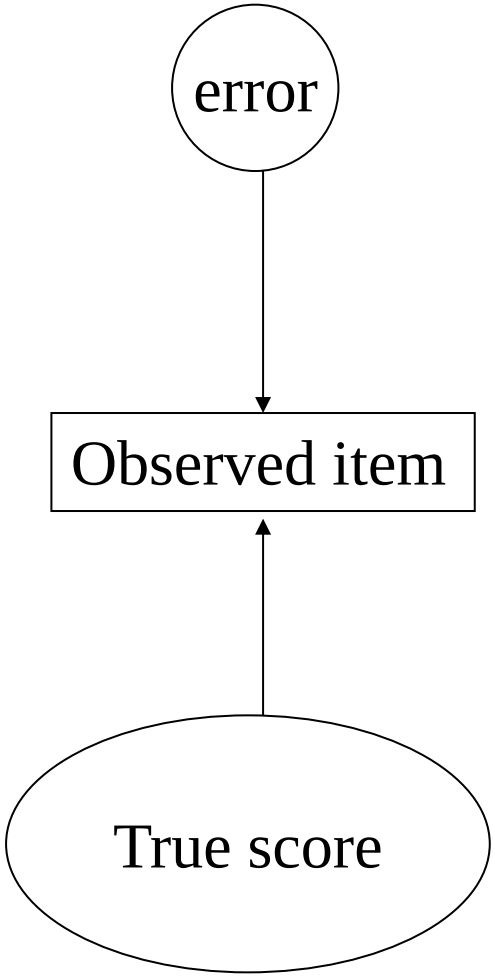
$$X = t + e$$

Observed item

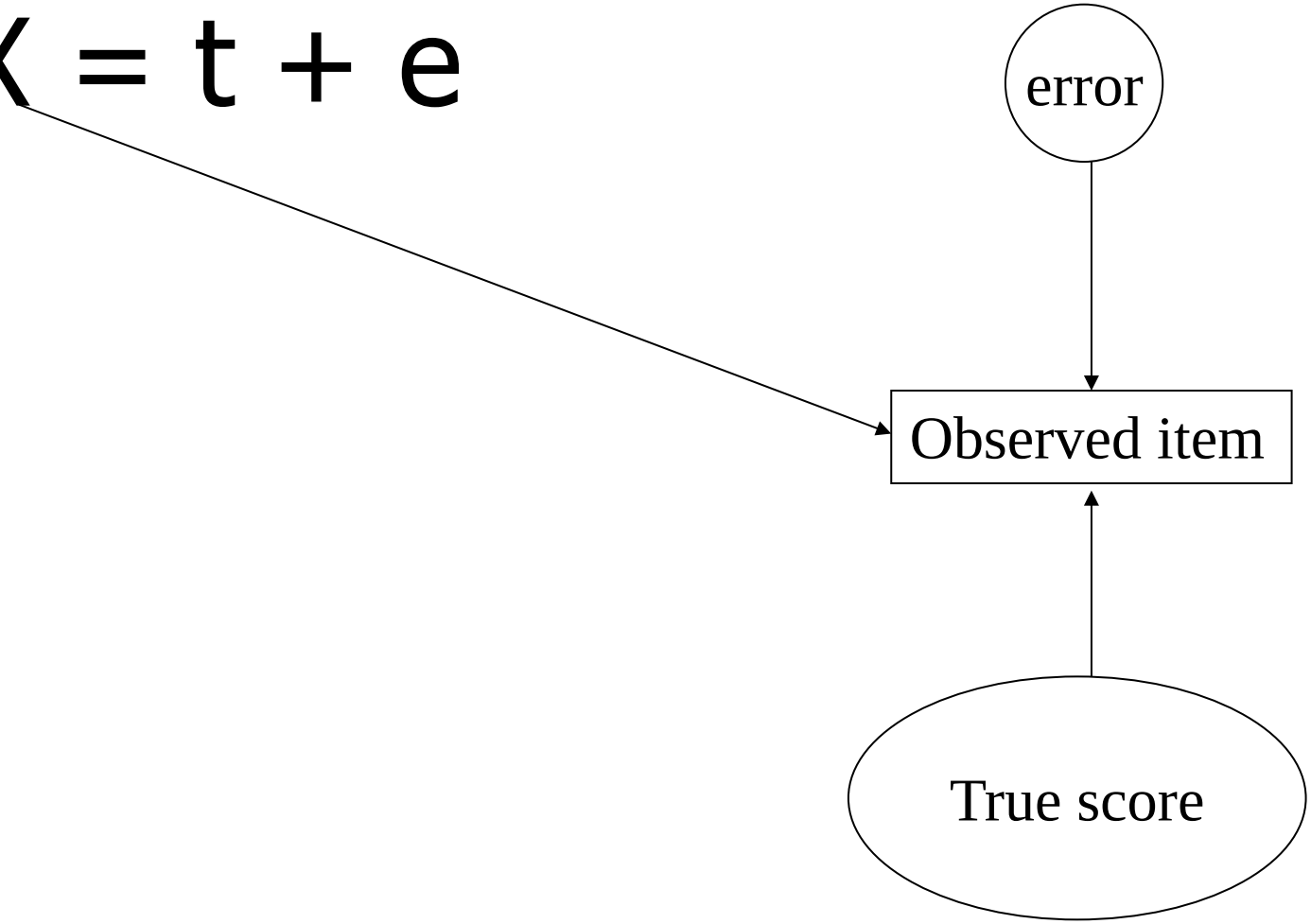
$$X = t + e$$



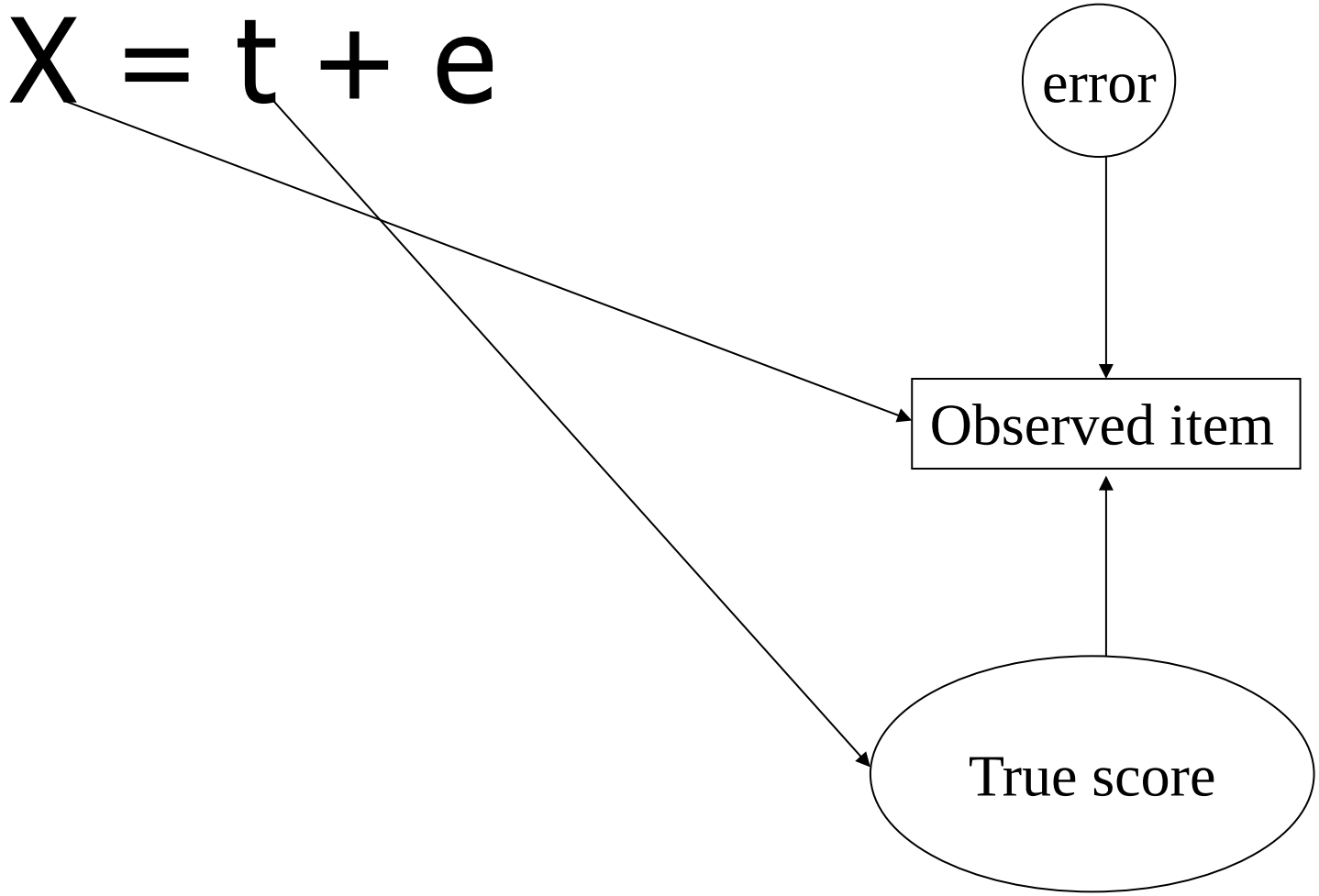
$$X = t + e$$

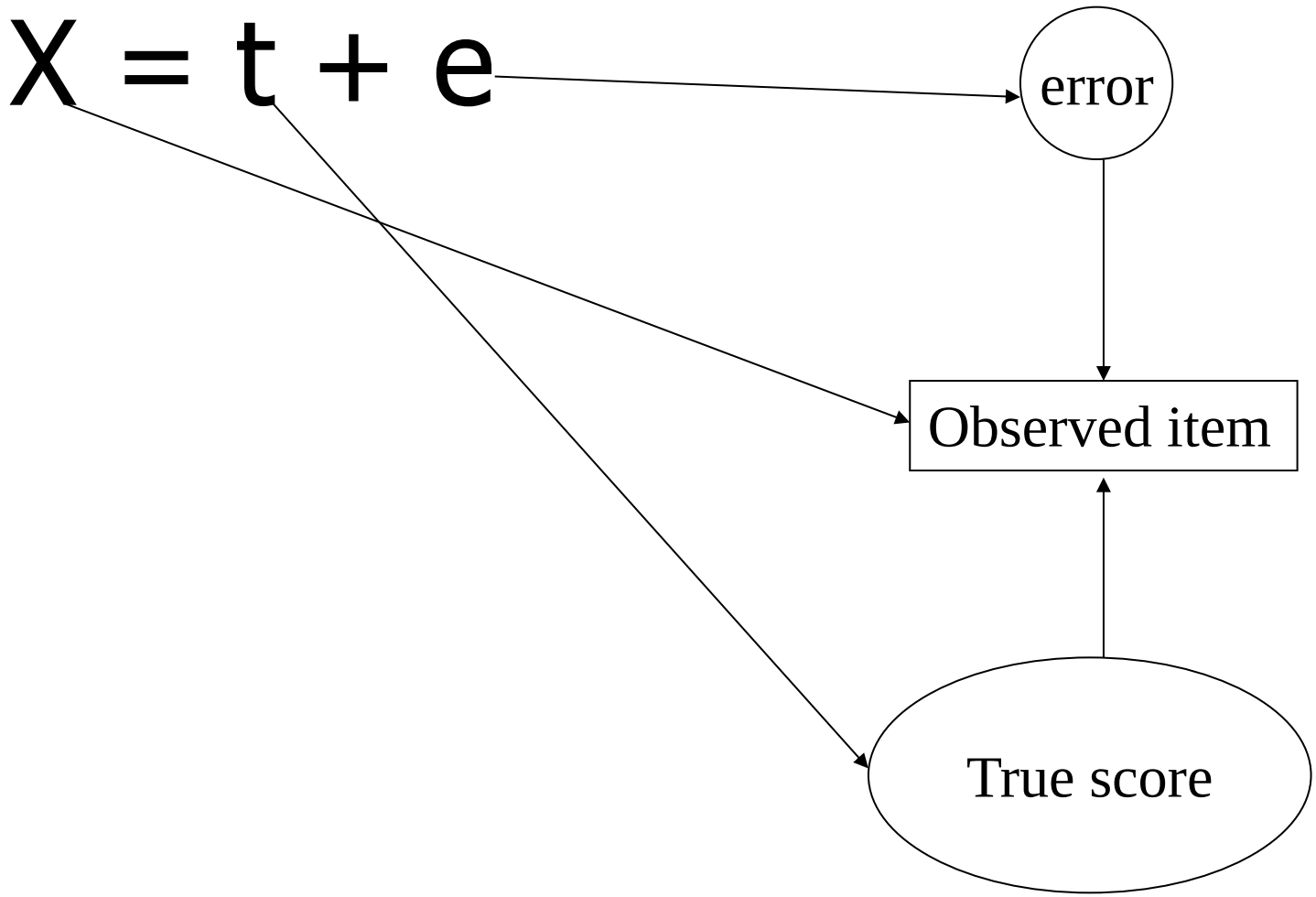


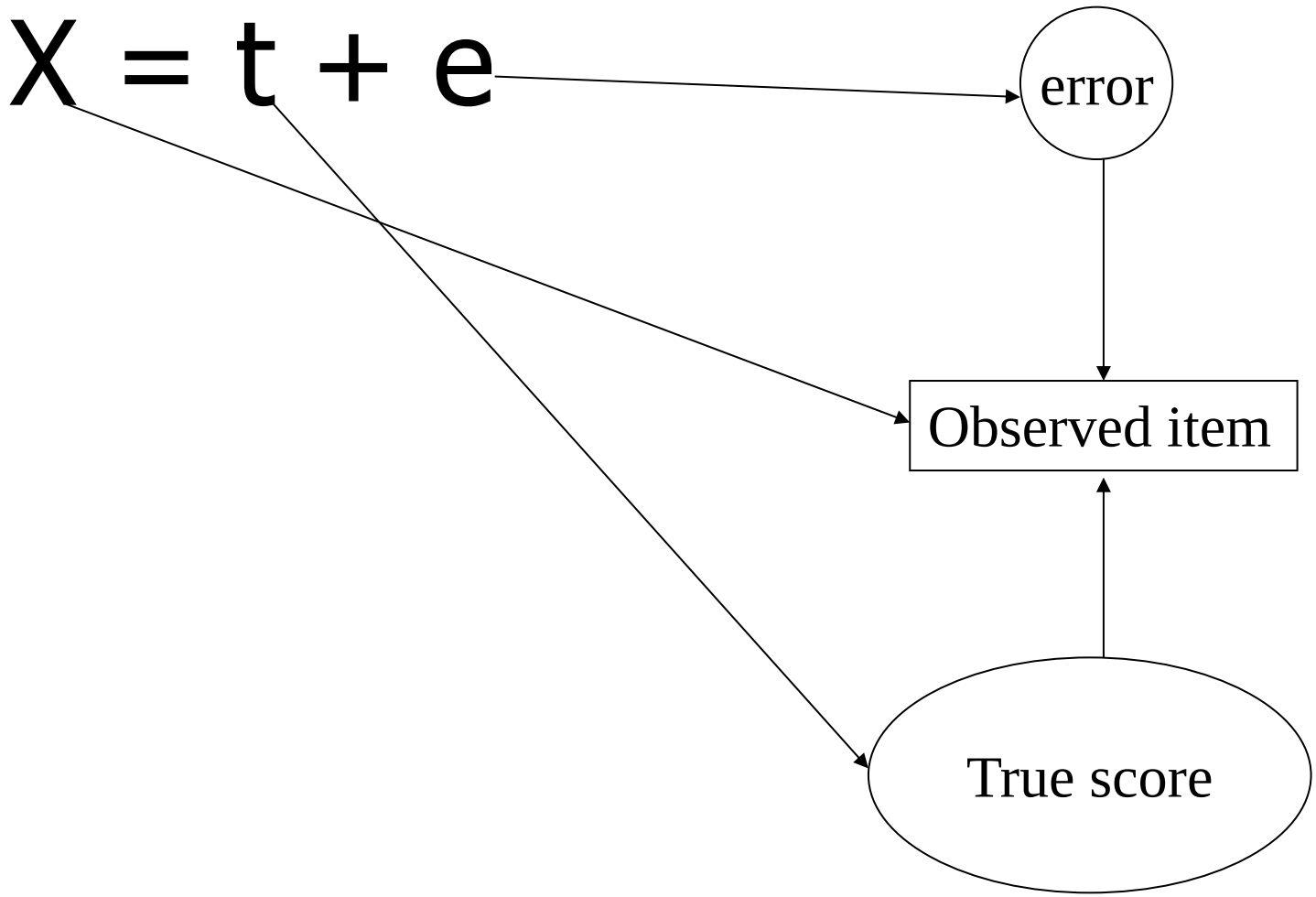
$X = t + e$



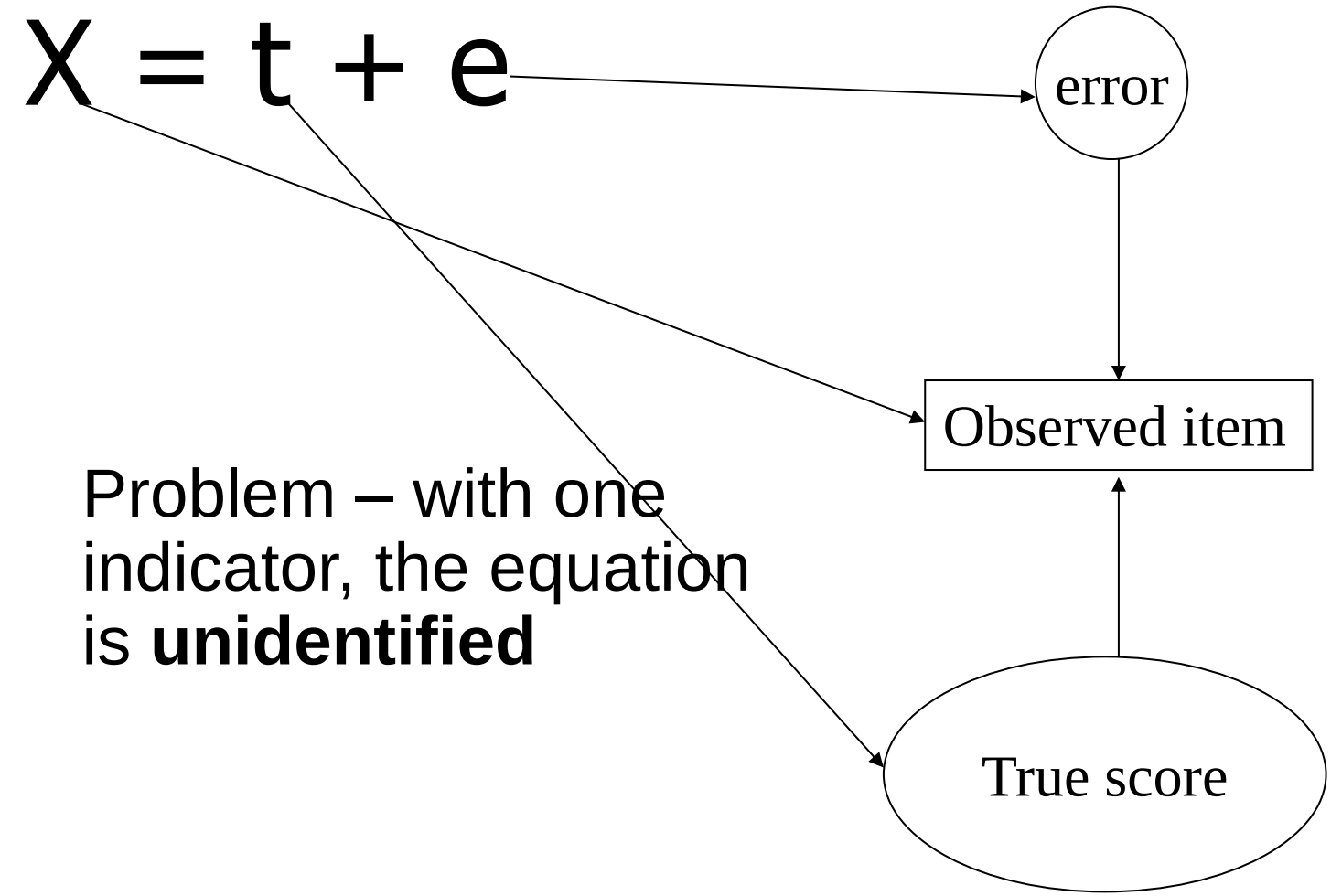
$X = t + e$




X = t + e
error
Observed item
True score

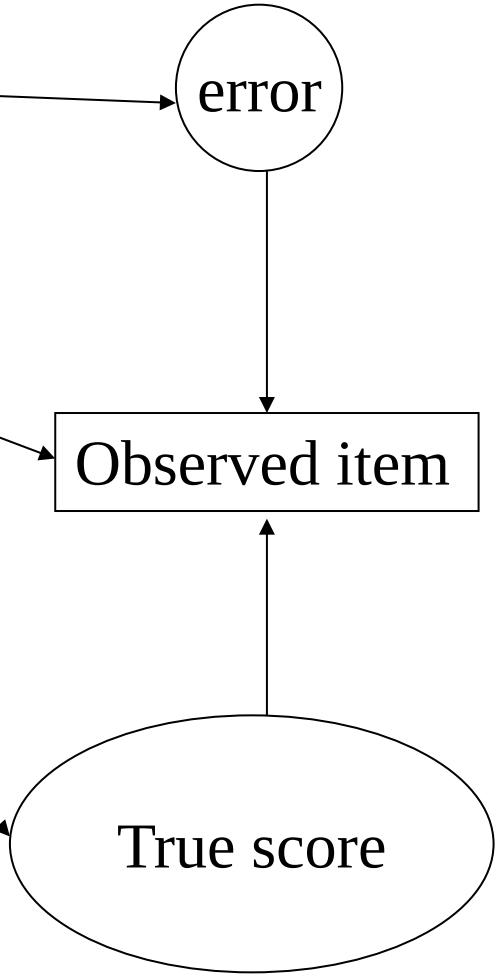

X = t + e
error
Observed item
True score

$X = t + e$



Problem – with one indicator, the equation is **unidentified**

$X = t + e$



Problem – with one indicator, the equation is **unidentified**

We can't separate true score and error

Multiple Indicator Latent Variables

Multiple Indicator Latent Variables

- To identify t & e components we need multiple indicators of the latent variable

Multiple Indicator Latent Variables

- To identify t & e components we need multiple indicators of the latent variable
- With multiple indicators we can use a latent variable model to partition variance

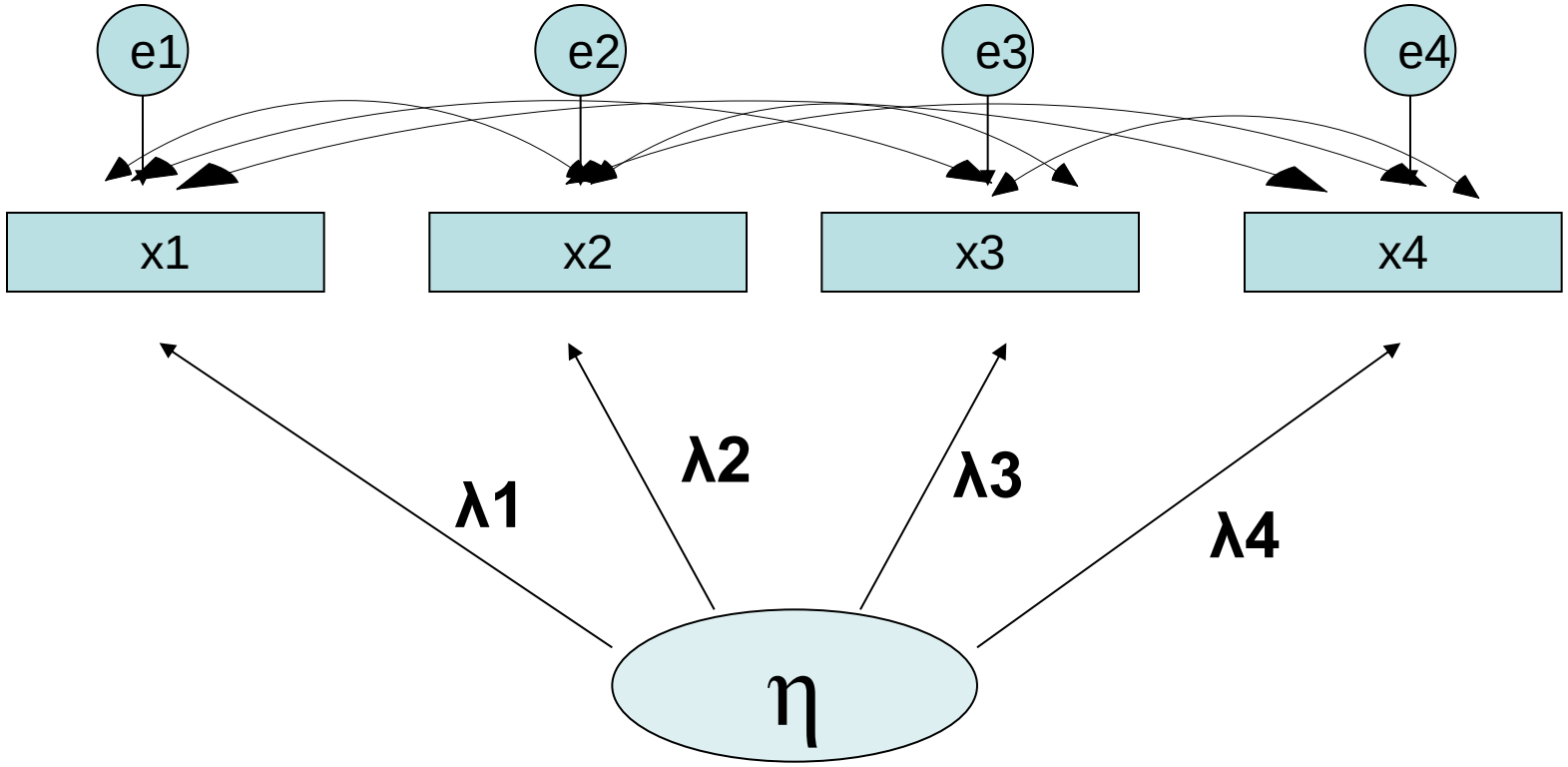
Multiple Indicator Latent Variables

- To identify t & e components we need multiple indicators of the latent variable
- With multiple indicators we can use a latent variable model to partition variance
- e.g. principal components analysis transforms correlated variables into uncorrelated components

Multiple Indicator Latent Variables

- To identify t & e components we need multiple indicators of the latent variable
- With multiple indicators we can use a latent variable model to partition variance
- e.g. principal components analysis transforms correlated variables into uncorrelated components
- We can then use a reduced set of components to summarise the observed associations

A Common Factor Model



λ = Factor loadings = correlation between factor & indicator

Benefits of Latent Variables

Benefits of Latent Variables

- Most social concepts are complex and multi-faceted

Benefits of Latent Variables

- Most social concepts are complex and multi-faceted
- Using single measures will not adequately cover the full conceptual map

Benefits of Latent Variables

- Most social concepts are complex and multi-faceted
- Using single measures will not adequately cover the full conceptual map
- Removes/reduces random error in measured construct

Benefits of Latent Variables

- Most social concepts are complex and multi-faceted
- Using single measures will not adequately cover the full conceptual map
- Removes/reduces random error in measured construct
- Random error in dependent variables -> estimates unbiased but less precise

Benefits of Latent Variables

- Most social concepts are complex and multi-faceted
- Using single measures will not adequately cover the full conceptual map
- Removes/reduces random error in measured construct
- Random error in dependent variables -> estimates unbiased but less precise
- Random error in independent variables -> attenuates regression coefficients toward zero

Remember

SEM can be thought of as

Path Analysis

using

Latent Variables

Remember

SEM can be thought of as

Path Analysis

using

Latent Variables

We now know about latent variables, what about path analysis?

Path Analysis

Path Analysis

- The diagrammatic representation of a theoretical model using standardised notation

Path Analysis

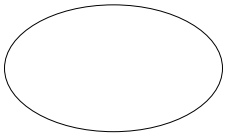
- The diagrammatic representation of a theoretical model using standardised notation
- Regression equations specified between measured variables

Path Analysis

- The diagrammatic representation of a theoretical model using standardised notation
- Regression equations specified between measured variables
- ‘Effects’ of predictor variables on criterion/dependent variables can be:
  - Direct
  - Indirect
  - Total

Path Diagram notation

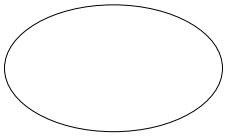
Path Diagram notation



Measured latent variable

Observed / manifest variable

Path Diagram notation

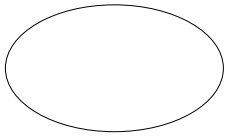


Measured latent variable

Observed / manifest variable

Error variance / disturbance term

Path Diagram notation



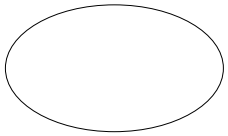
Measured latent variable

Observed / manifest variable

Error variance / disturbance term

Covariance / non-directional path

Path Diagram notation



Measured latent variable

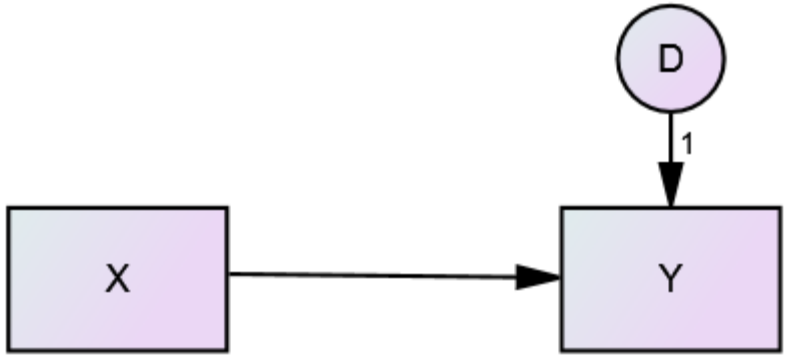
Observed / manifest variable

Error variance / disturbance term

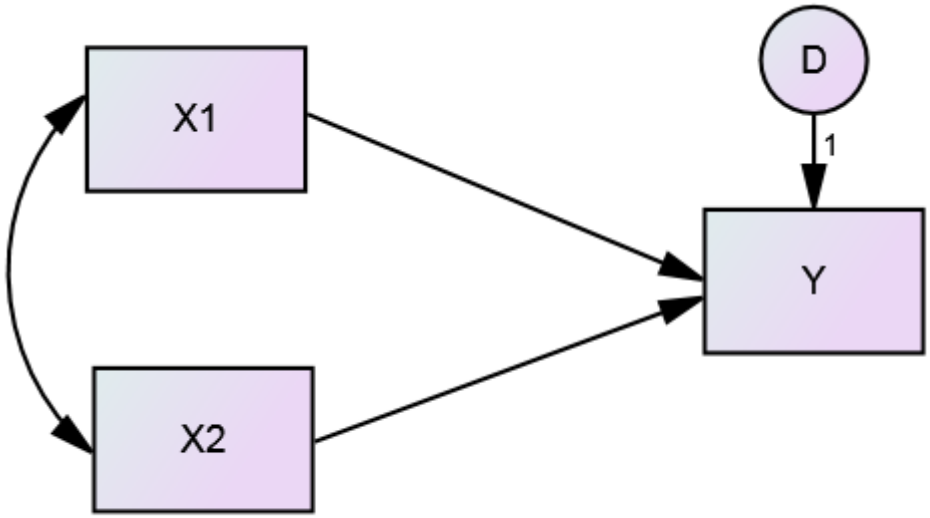
Covariance / non-directional path

Regression / directional path

PD1: Single Cause

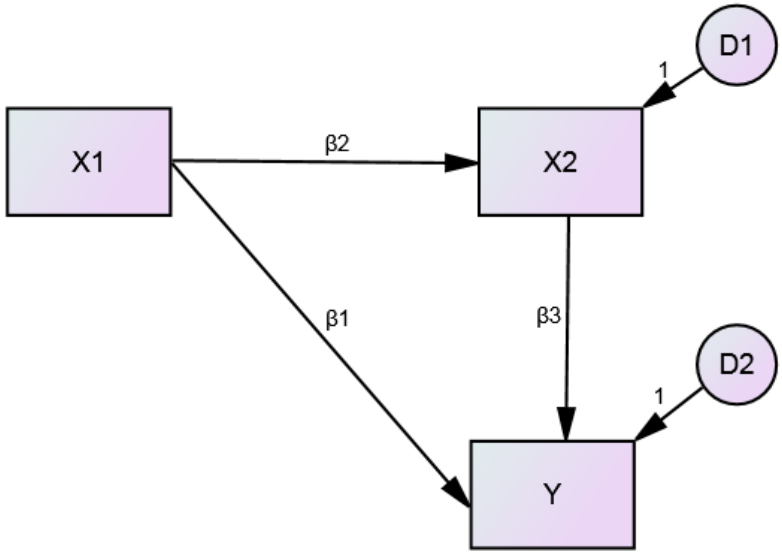


Two correlated causes

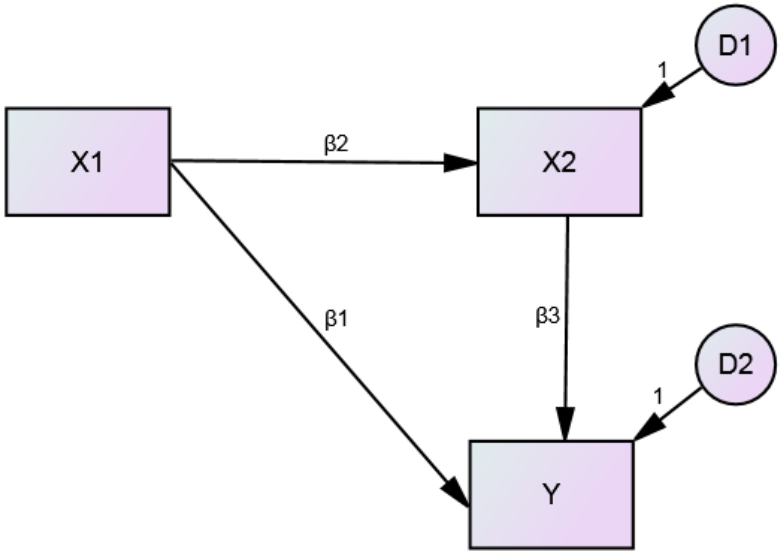


Indirect Effect

Indirect Effect

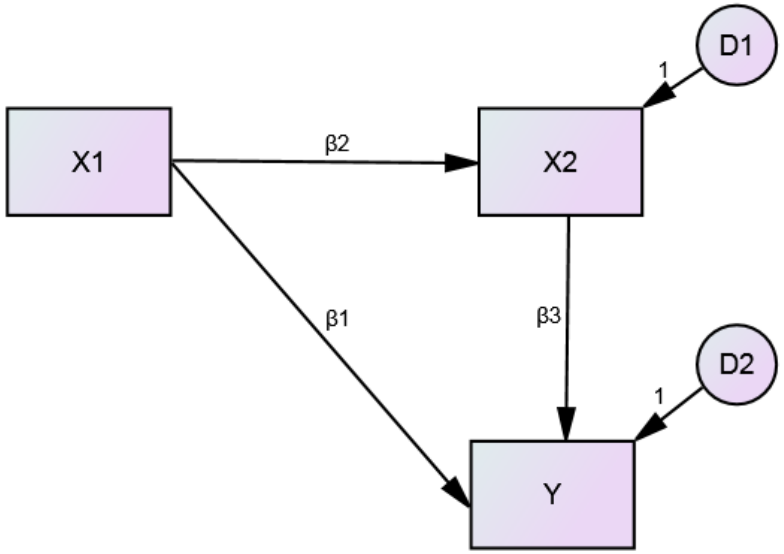


Indirect Effect



β1=direct effect of X1 on Y

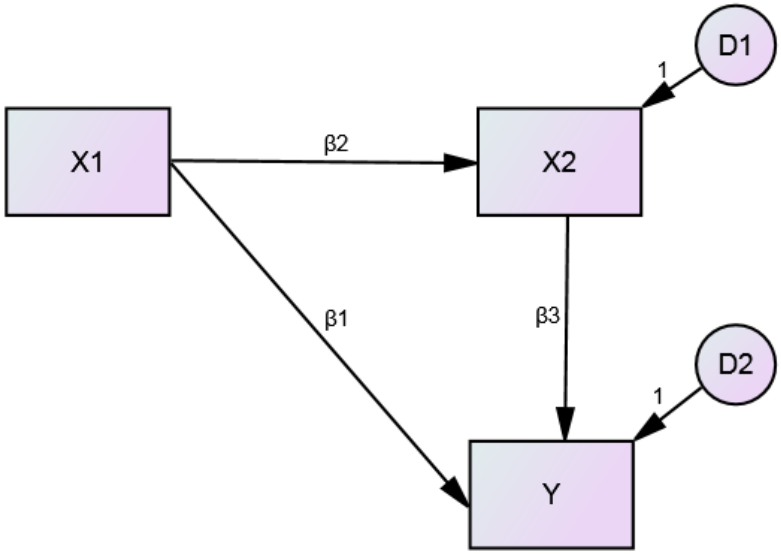
Indirect Effect



β1=direct effect of X1 on Y

β2=direct effect of X1 on X2

Indirect Effect

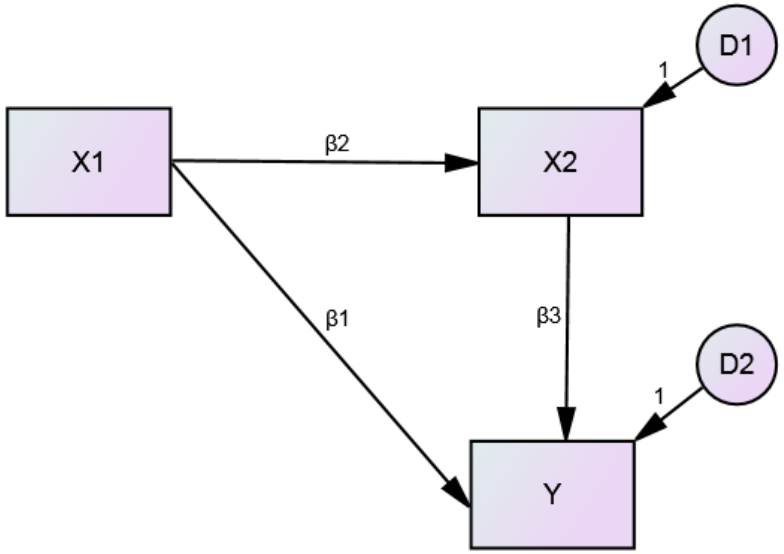


β1=direct effect of X1 on Y

β2=direct effect of X1 on X2

β3=direct effect of X2 on Y

Indirect Effect



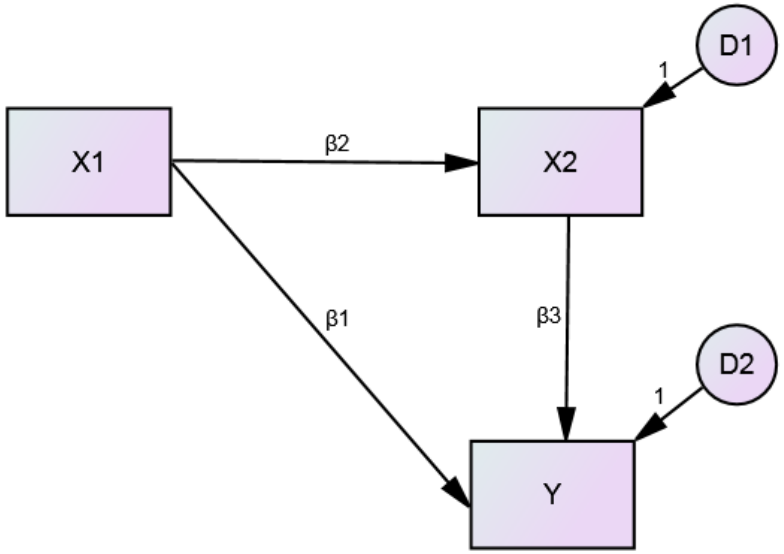
β1=direct effect of X1 on Y

β2=direct effect of X1 on X2

β3=direct effect of X2 on Y

β2*β3=indirect effect of X1 on Y

Indirect Effect



β1=direct effect of X1 on Y

β2=direct effect of X1 on X2

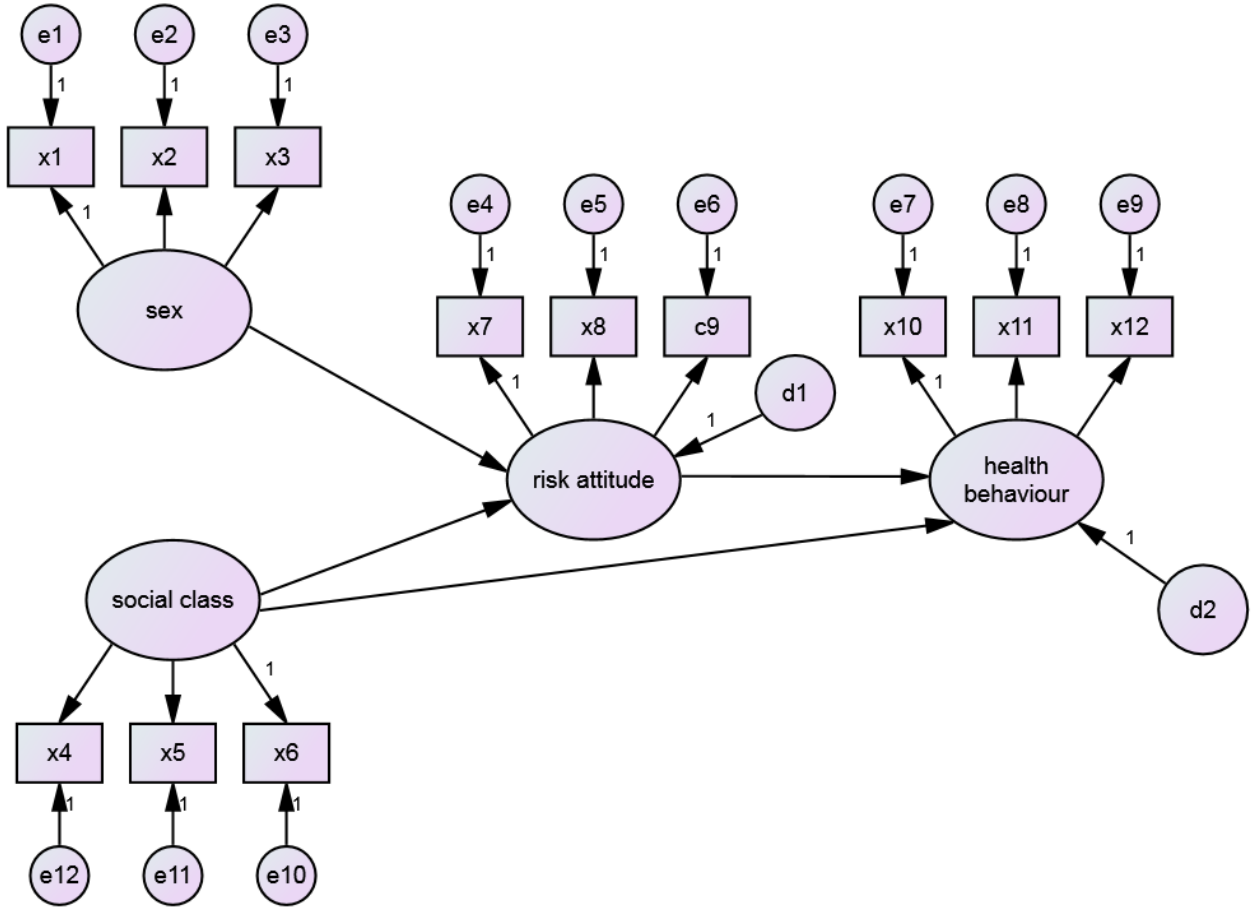
β3=direct effect of X2 on Y

β2*β3=indirect effect of X1 on Y

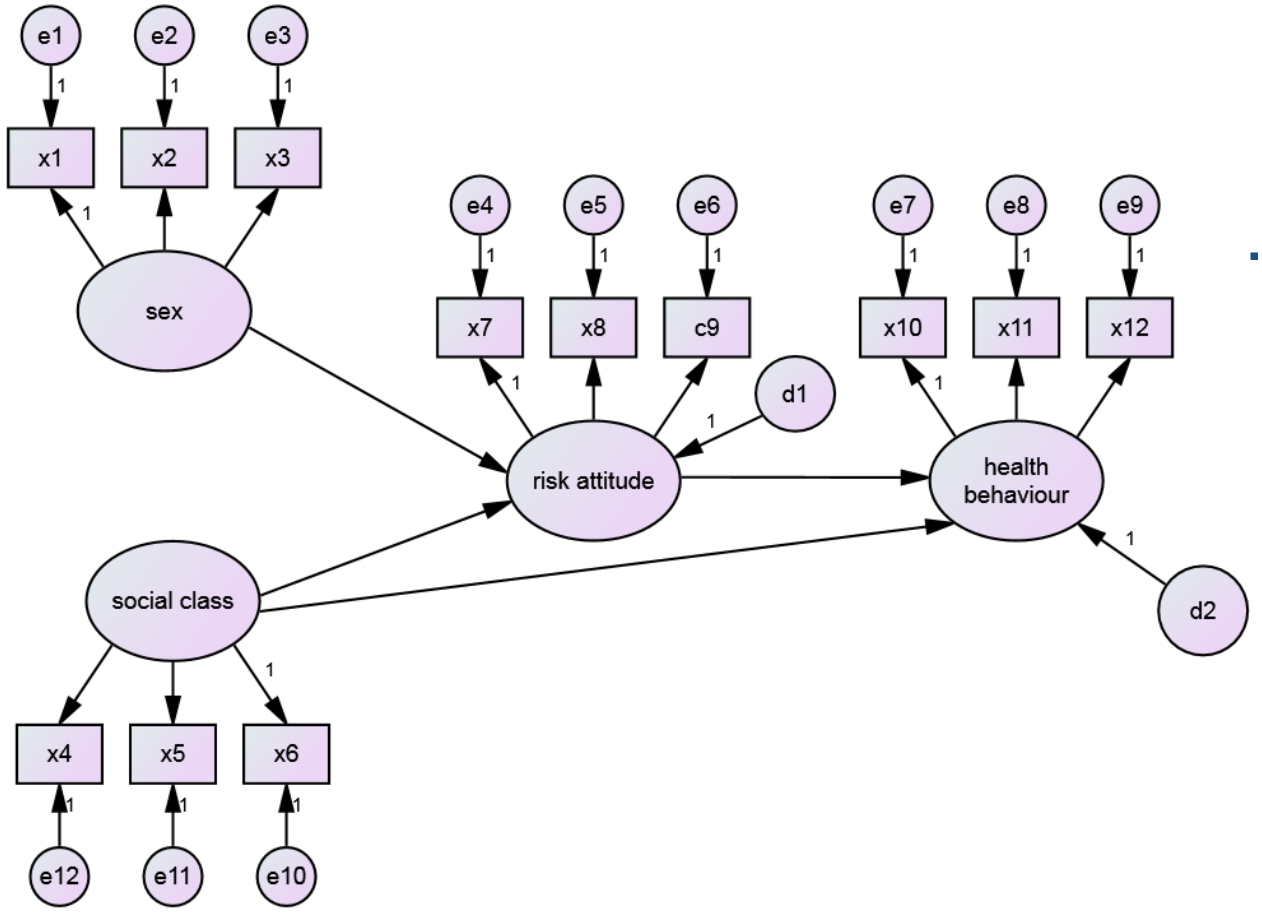
β1+(β2*β3)=total effect of X1 on Y

So a path diagram with latent variables…

So a path diagram with latent variables…



So a path diagram with latent variables…



…is a SEM

NCRM National Centre for Research Methods

For more information contact

ncrm.ac.uk