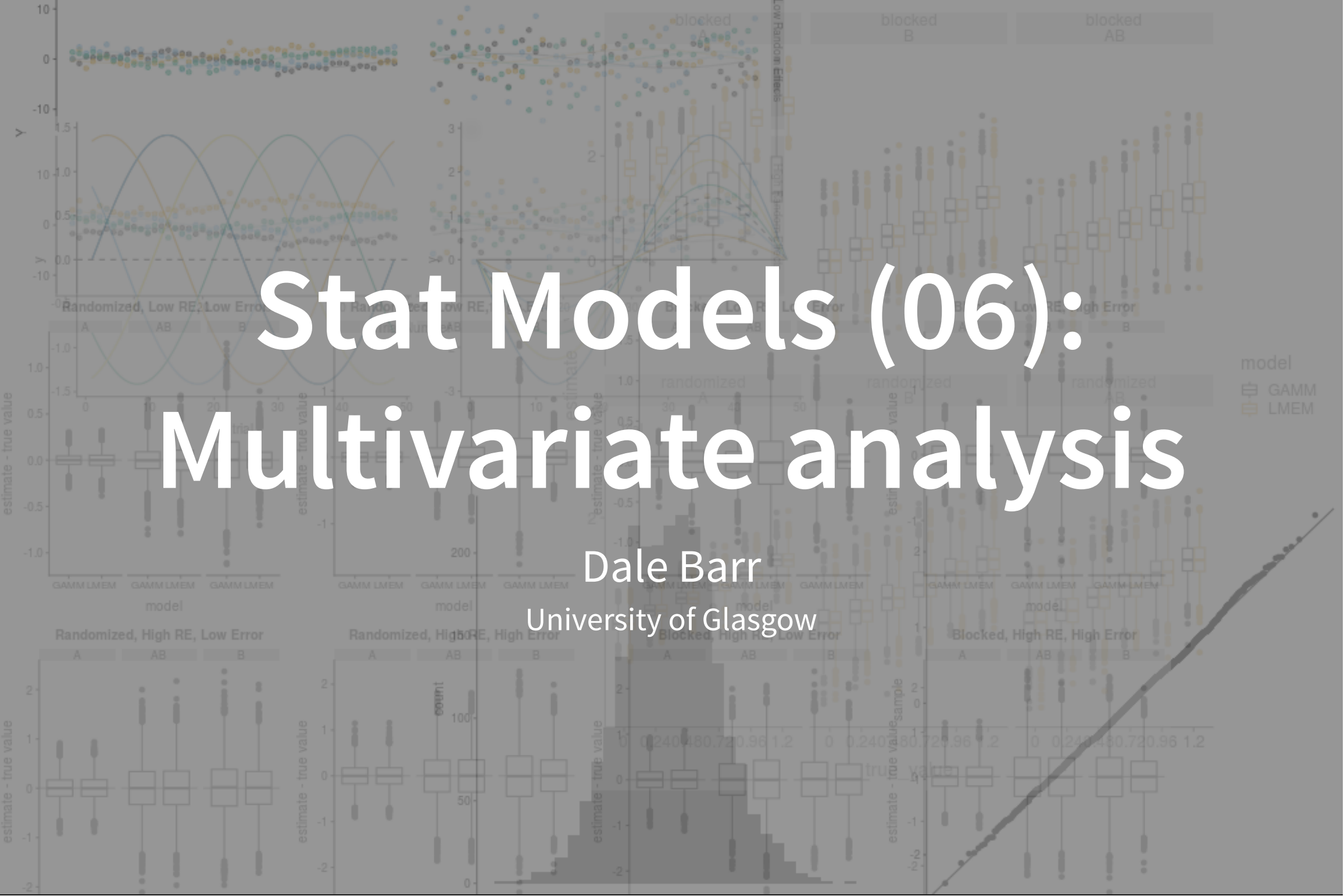


Stat Models (06): Multivariate analysis

Dale Barr

University of Glasgow

part II

lecture	topic
6	introduction to multivariate analysis
7	path analysis
8	mediation models
9	confirmatory factor analysis
10	structural equation modeling

multivariate analysis

- more than one response variable (or DV)
- focus on causal relationships / patterns of associations between variables
- modeling framework: Structural Equation Models (SEM)
  - structural model: path / mediation analysis
  - measurement model: CFA, full SEM

the SEM framework

- brings together regression & psychometrics
  - making it possible to take measurement error/ reliability into account
- allows for estimation of indirect and direct (causal) effects

Also Known As (AKA)

- covariance structure analysis/modeling
- Analysis of Moment Structures (AMOS)
- Linear Structural Relations (LISREL)

other families of SEM:

- variance-based SEM / partial least squares path Modeling (PLS-PM), common in marketing/organizational research
- structural causal model (SCM) or nonparametric SEM
  - directed graphs (DAG / DCG)

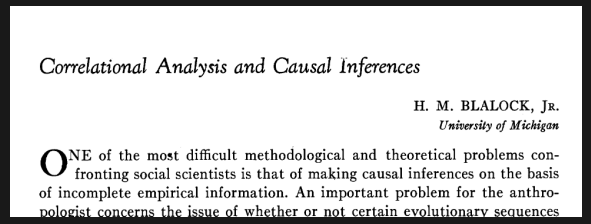
what SEM does

- analyst specifies a model based on domain knowledge
- estimate model parameters to minimize the difference between:
  - sample covariance matrix (observed)
  - covariance matrix implied by the model

can estimate means too, but usually focus is on covariances

history of SEM

- Sewell Wright (geneticist) method of path coefficients (1934)
- introduced to social sciences by Blalock (1961) and others
- led to integration of regression with factor-analytic techniques (JWK model) Karl Jöreskog, J. W. Keesling, D. Wiley
- first publicly available computer program was LISREL III by Jöreskog & Sörbom (1976)

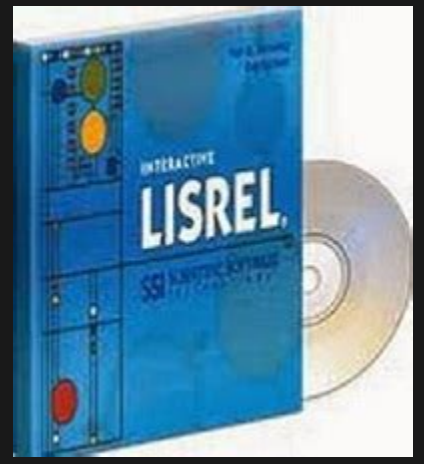


Correlational Analysis and Causal Inferences

H. M. BLALOCK, JR.
University of Michigan

ONE of the most difficult methodological and theoretical problems confronting social scientists is that of making causal inferences on the basis of incomplete empirical information. An important problem for the anthropologist concerns the issue of whether or not certain evolutionary sequences

Blalock (1961)



Klein (2023), Principles and Practice of Structural Equation Modeling, 5th Ed. New York: The Guilford Press.

where SEM is mostly used

- industrial/organisational psychology
- marketing
- educational psychology
- health sciences
- psychological assessment
- neuropsychology
- political sciences
- economics

mostly observational / nonexperimental studies, but also, experimental / quasi-experimental studies (Breitsohl, 2019)

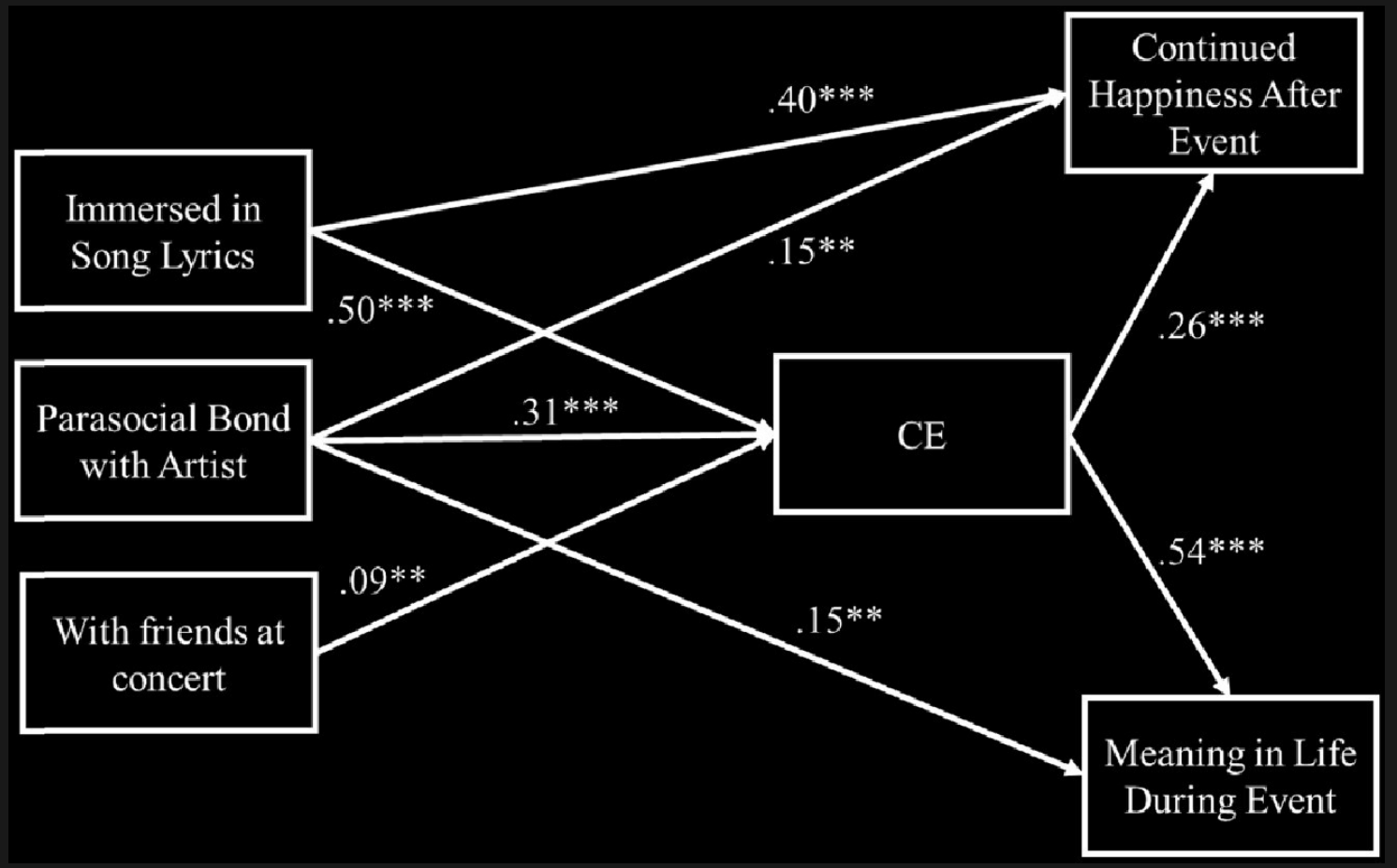
disadvantages / challenges

- NOT a tool for causal discovery
- it is a “large-sample technique”
  - N:q rule, ratio of number of cases (N) to number of model parameters needing estimates (q). 20:1 is recommended (Jackson 2003)
  - most published SEM studies are probably based on samples that are too small (Loehlin & Beaujean, 2017)
  - most do not adequately report:
    - rationale for SEM, for sample size, missing data, reliability of variables, correspondence between model & data (Zhang et al. 2021)
- assumes linearity / multivariate normality

software

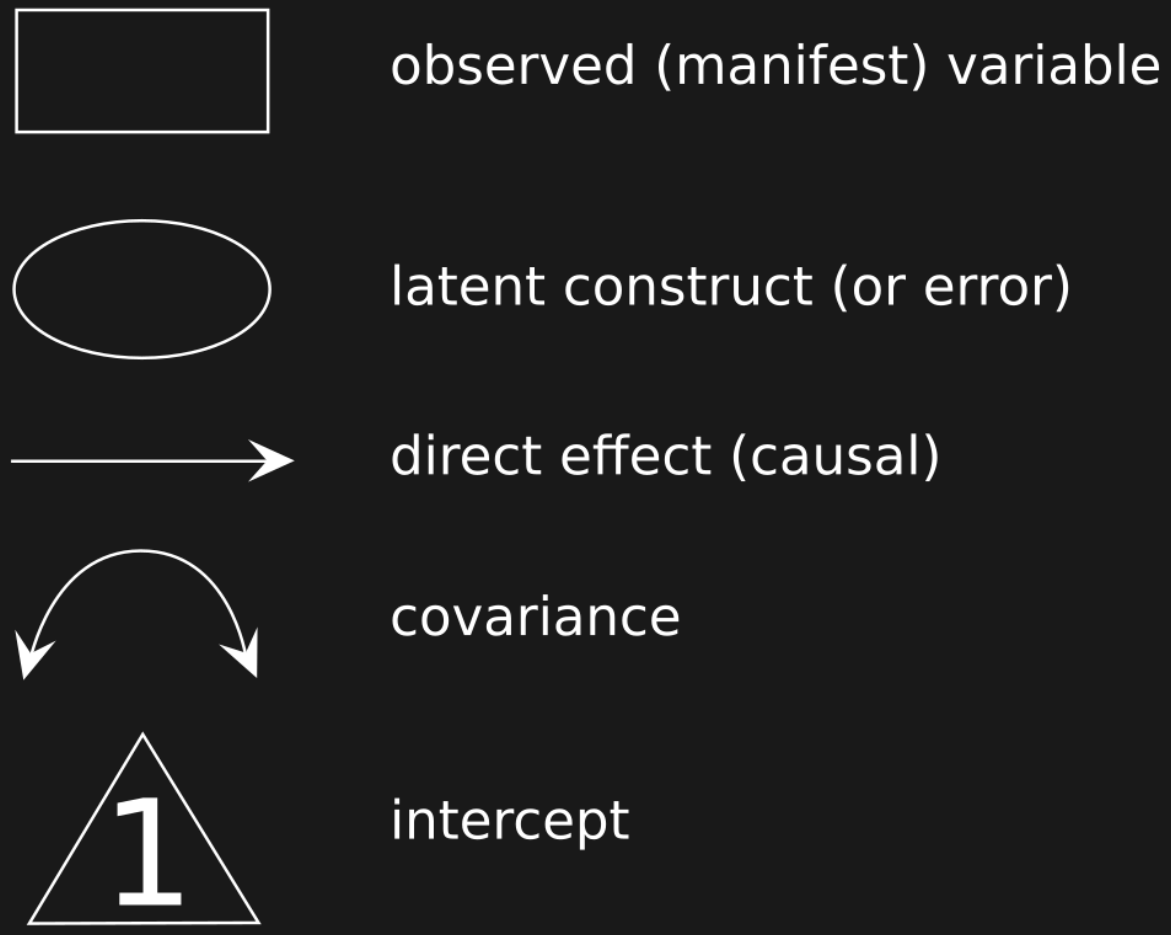
- open-source
  - R: `lavaan`, `semTools`, `OpenMx`
  - JASP
- commercial
  - Amos, LISREL, Mplus

path diagrams



Koefler et al. (2024), Let the music play: Live music fosters collective effervescence and leads to lasting positive outcomes. *Personality and Social Psychology Bulletin*.

model diagram syntax



“McArdle-McDonald RAM syntax”

a simple regression model

manifest vs latent

in the social sciences, we often can only measure ‘proxies’ for the true theoretical constructs that we are interested in (e.g., happiness, job satisfaction, self-esteem, statistics anxiety, social capital)

importance of reliability

- incremental validity: does a predictor have an effect on some response *after controlling for a second predictor?*
- multiple regression assumes perfect score reliability!

Type I error rates of significance tests are surprisingly high even in large samples (N = 300) with moderate levels of predictor reliability (.8) in analyses with 2 predictors

“incremental validity” better supported by SEM

Westfall & Yarkoni (2016), Statistically Controlling for Confounding Constructs Is Harder than You Think.

ice cream & swimming deaths

SEM analysis workflow

1. specify the model
2. evaluate whether the model is identified
3. select measures (operationalize) and collect data
4. analyze the model
    - if fit is poor, respecify (step 5) if justified by theory; if not, retain no model
    - if model retained, interpret parameter estimates
    - consider equivalent / near-equivalent models
5. respecify
6. report the results

Klein (2023), Principles and Practice of Structural Equation Modeling, 5th Ed. New York: The Guilford Press.

model specification

represent your hypotheses as equations or (more commonly) as a path diagram

- endogenous variable: one or more causal paths leading into it
- intervening variable: one or more causal path leading in and one or more leading out
- exogenous variable: one or more causal path leading out, none leading in

guided by domain knowledge, not by the analysis!

model identification

- analyzing a SEM requires deriving estimates for parameters in a set of *simultaneous equations*
- a parameter is ‘identified’ if a unique value can be found given the model and the data matrix

Term	Description
under-identified	at least one unidentified parameter
just-identified or saturated	data complexity = model complexity
over-identified	model simpler than data