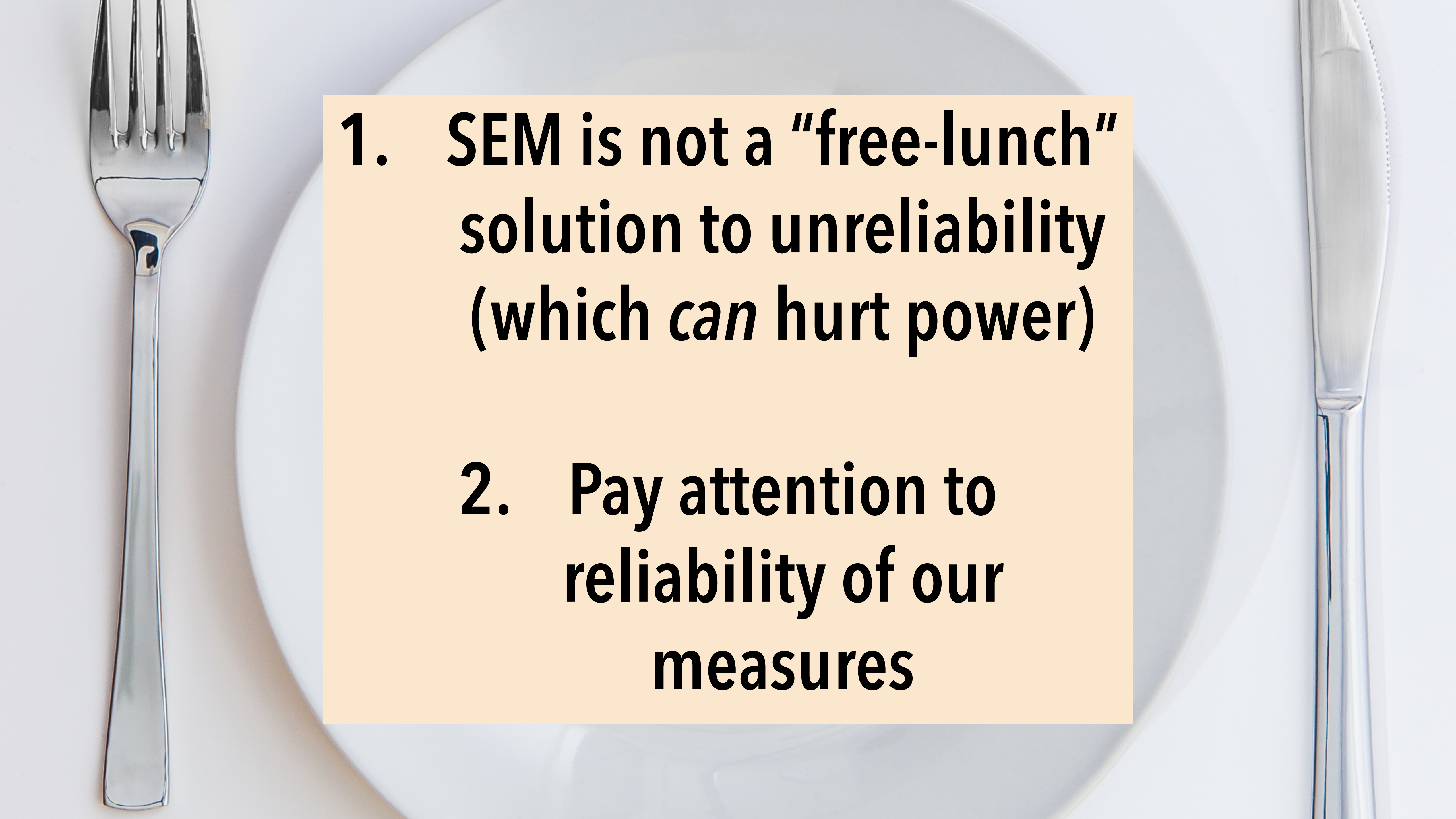
Connecting Unreliable Measurement to Statistical Power in Structural Equation Modeling (SEM)

Y. ANDRE WANG & MIJKE RHEMTULLA

APS 2018

UCDAVIS
DEPARTMENT OF PSYCHOLOGY


1. SEM is not a "free-lunch" solution to unreliability (which *can* hurt power)
2. Pay attention to reliability of our measures

Power is important to psychology

Highly powered studies are more likely to…

- Detect true effects
- Produce replicable results
- Buffer against false positives in the literature

Valuable to (cumulative) psychological science

Journals call for greater emphasis on power

However, power remains low

Button et al., 2013; Gervais et al., 2015; Marszalek et al., 2011; Maxwell, 2004; Schimmack, 2012; Vazire, 2015

Unreliable measures can compromise power

Measurement error in variables can introduce “noise” and reduces statistical power

Many established psychological measures have substantial measurement error

- Reaction time based cognitive measures (e.g., Stroop) often have low ICCs (e.g., .4–.7)

Unreliable measures can pose a challenge to running powerful studies

Cole & Preacher, 2013; Spearman, 1910; Hedge, Powell, & Sumner, 2017

SEM: A solution to measurement error?

Structural equation modeling (SEM) is often touted as a solution to measurement error

Can correct for measurement error by estimating unreliability at the measurement level

This is a big draw for researchers!

Bollen, 1989; Hoyle, 1995; Oberski & Satorra, 2013

Examples: SEM & Measurement Error

Testing relations among implicit racial attitude measures

> *"…**low reliability (high measurement error)** need not be a threat to construct validity…analyses that utilize latent variable models…circumvent this problem…**reliability does not constrain validity in latent variable analyses**."*

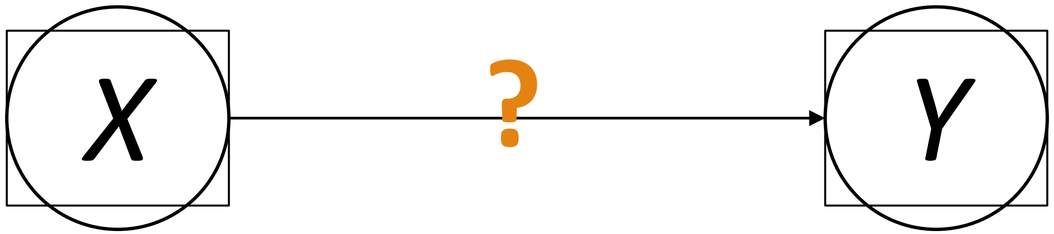
Assessing change in depression and anxiety in children

> *"Clearly, we need adequate methods for **taking measurement error into account** as we model change**…[this problem] can be handled with an SEM approach**."*

Cunningham, Banaji, & Preacher, 2001; Geiser, Eid, Nussbeck, Courvoisier, & Cole, 2010

SEM and Power

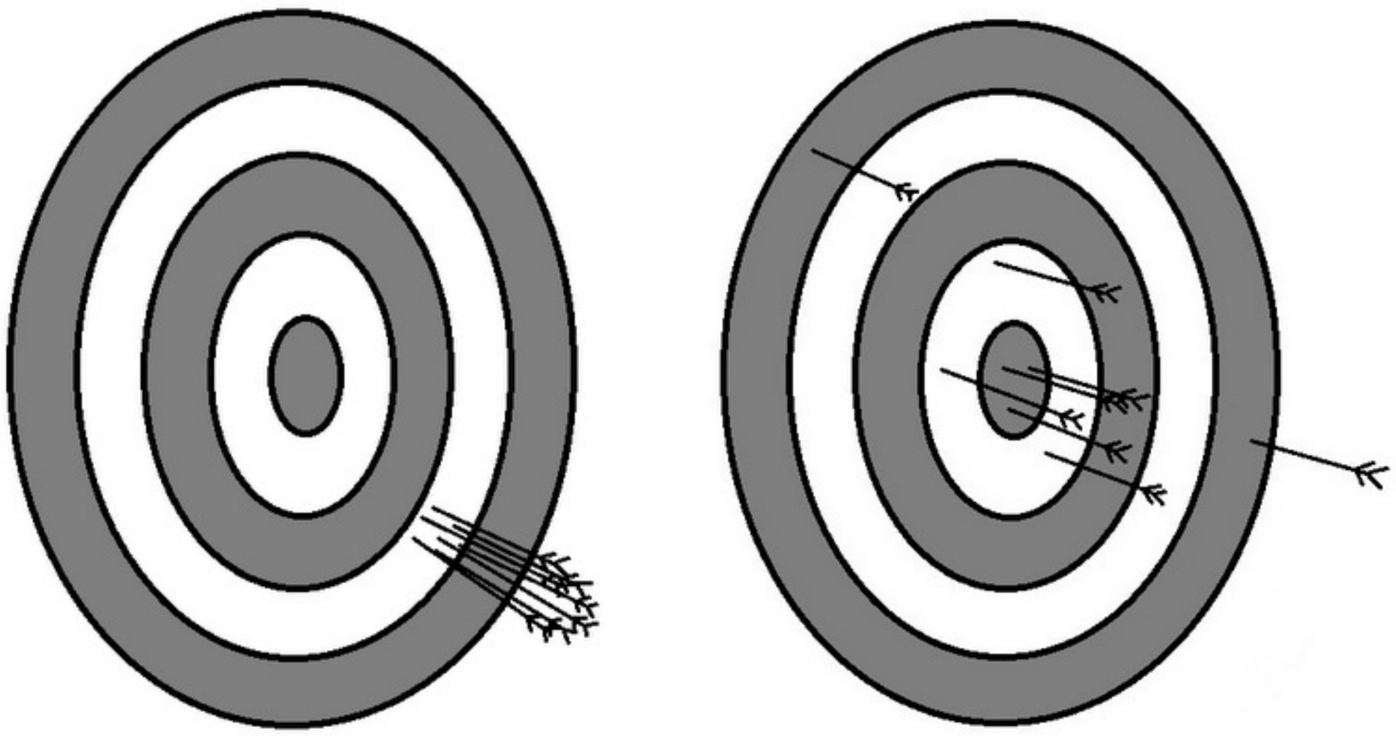
If SEM can account for measurement error, do we get more power with SEM?



SEM: More Accurate but Less Precise

SEM (vs. simple regression) is **more accurate but less precise**

Less precise = less power to detect structural parameters



The regression approach: Confidently inaccurate

The SEM approach: Accurate but uncertain

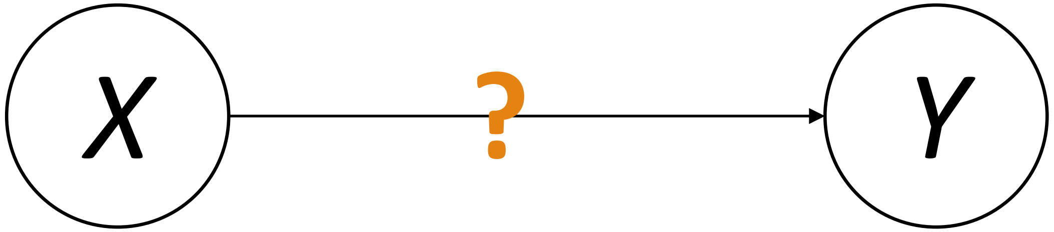
Hoyle & Kenny, 1999; Ledgerwood & Shrout, 2011;
Graphics reproduced with permission from https://ledgerwood.faculty.ucdavis.edu/methodological-interests/

Power to Detect Structural Parameters

What does this mean for power to detect structural parameters in practical terms?

- What impact does measurement error have?
- What other characteristics affect power, and to what extent?

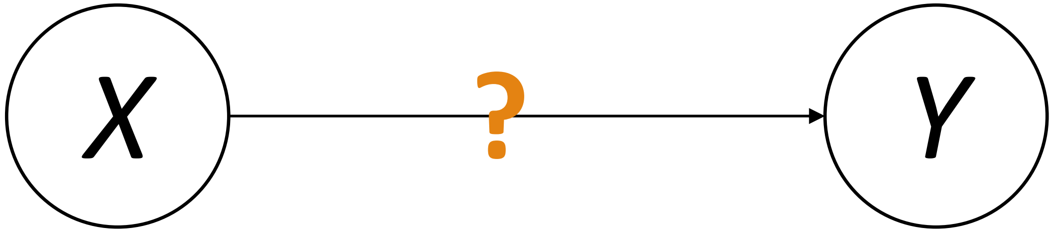
Simulation Study



We conducted a series of simulations

- Two-factor model
- Created known population models (true effect size of a parameter of interest is known and ≠ 0)
- Sample and analyzed datasets from the models to see how often we detect the regression parameter as significantly different from 0 (*power*)

We varied:



Sample size

- N = 50, 100, 200, 1000

Effect size

- β = .1, .2, .3, .4, .5 ($\cong r$ = .1–.5)

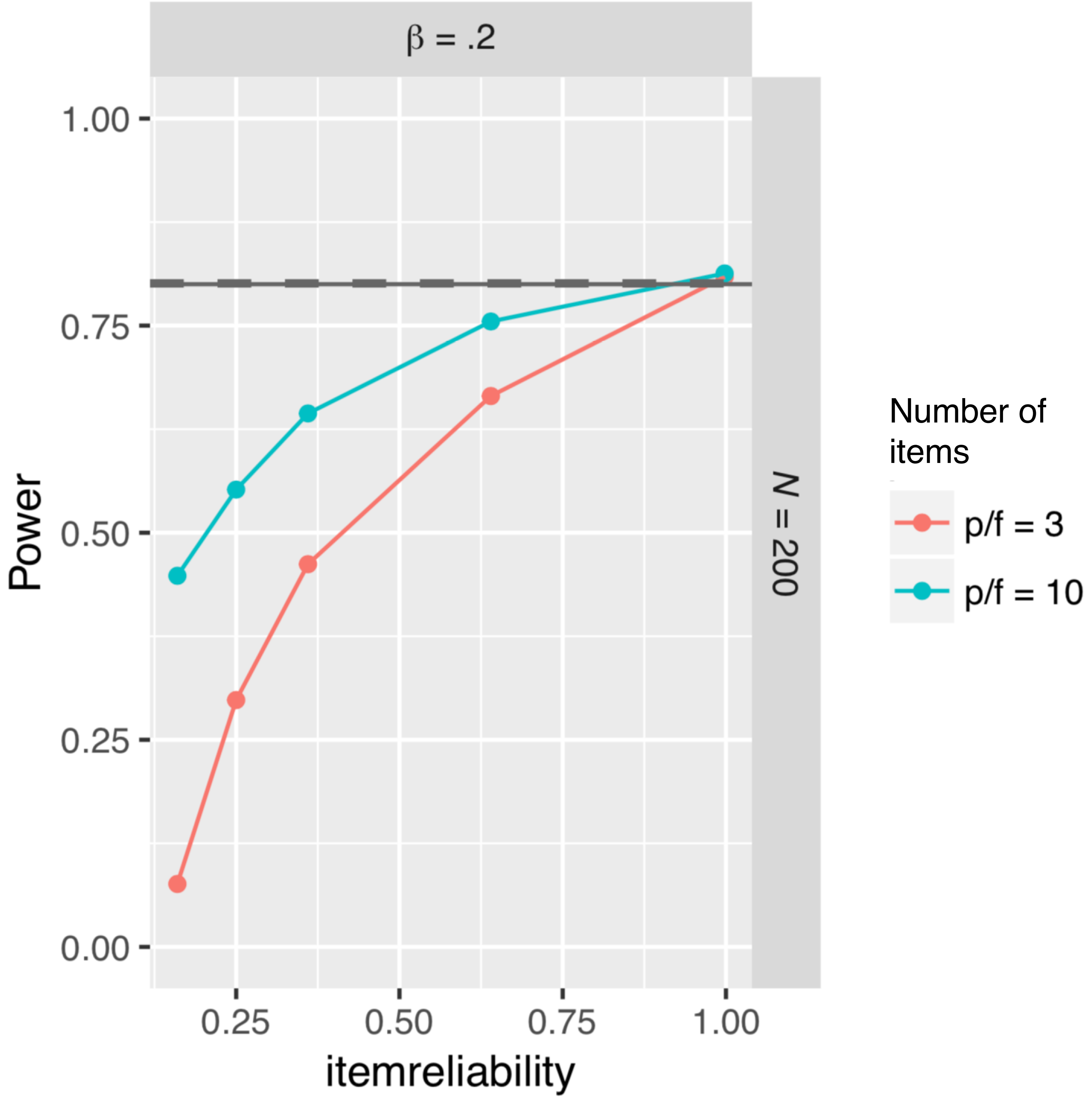
Number of items (per factor)

- p/f = 3, 10

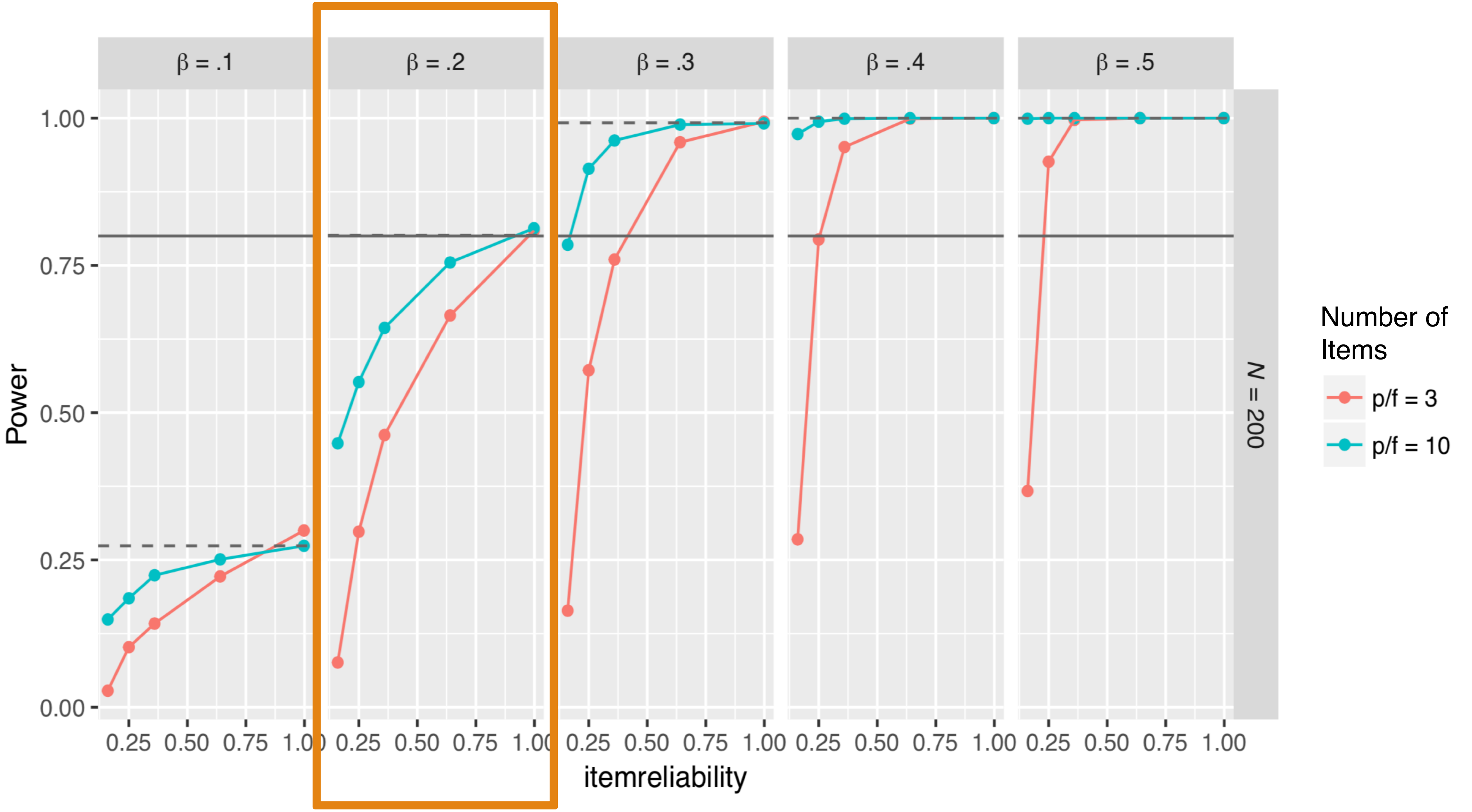
Factor loadings

- λ = .4, .5, .6, .8, .999 (R = .16, .25, .36, .64, .1)

_N_ = 200,
β = .2

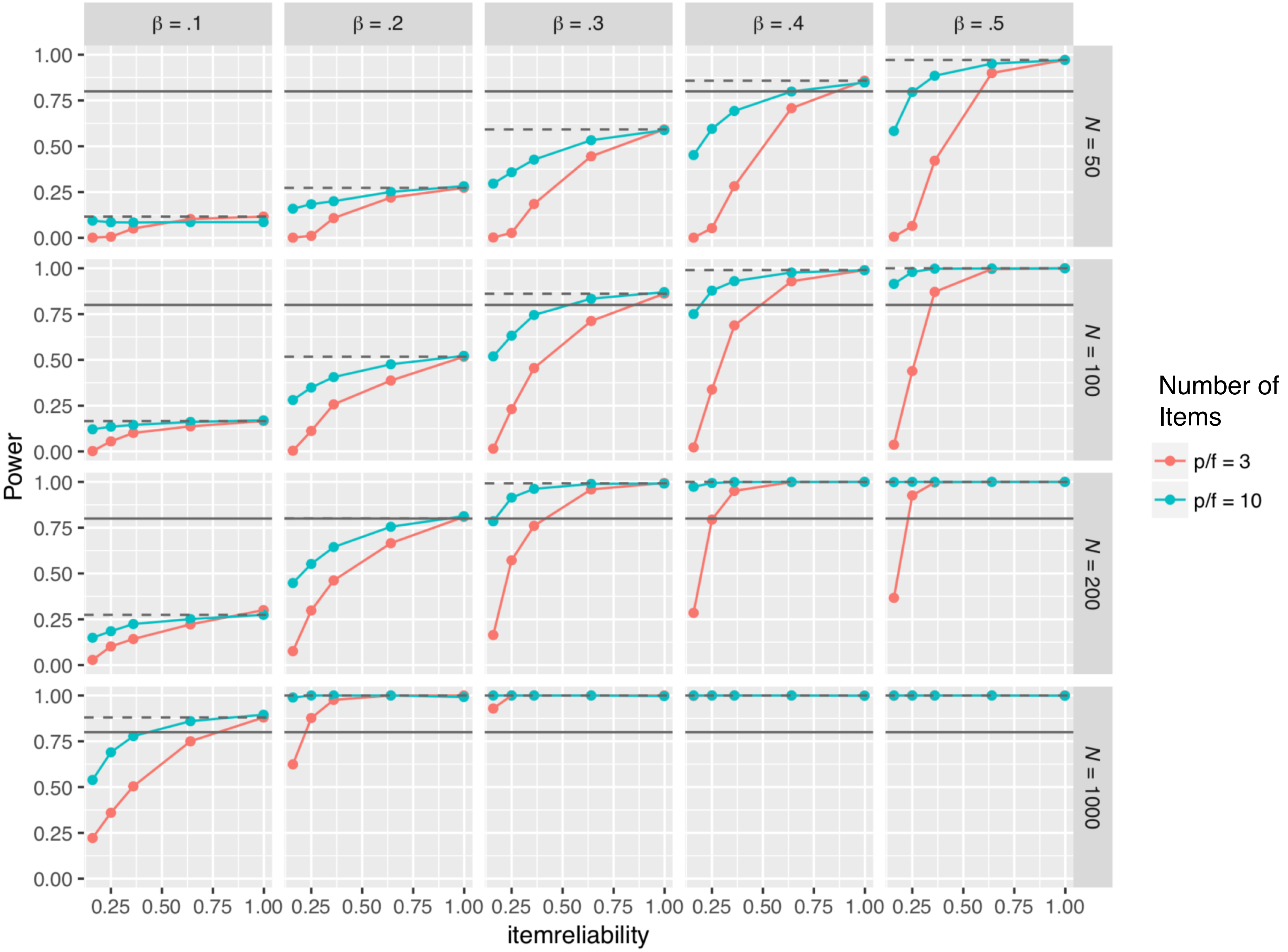


β = .2
N = 200
Power
1.00
0.75
0.50
0.25
0.00
0.25
0.50
0.75
1.00
itemreliability
Number of items
p/f = 3
p/f = 10



β = .1
β = .2
β = .3
β = .4
β = .5
Power
1.00
0.75
0.50
0.25
0.00
0.25 0.50 0.75 1.00
itemreliability
N = 200
Number of Items
p/f = 3
p/f = 10

N = 200



β = .1
β = .2
β = .3
β = .4
β = .5
N = 50
N = 100
N = 200
N = 1000
Power
itemreliability
Number of Items
p/f = 3
p/f = 10

Summary

SEM with unreliable measures can have lower power to detect non-zero (structural) parameters

- For unreliable measures (e.g., reaction time measures), this can spell trouble

Lower power hurts replicability of SEM findings

If you want to interpret specific effects in your model, take measurement reliability into account

- Likely means simulation-based power analysis

1. SEM is not a “free-lunch” solution to unreliability (which *can* hurt power)

2. Pay attention to reliability of our measures

If we are to take the science we do seriously,

then we need to take the measures we use to do science seriously.

Thank you! Questions?

Mijke Rhemtulla
UC Davis Attitude and Social Cognition Lab
UC Davis Social/Personality Psychology
UC Davis Quantitative Psychology

Compare with power to detect model misfit

In the MacCallum et al. framework, power to detect model misfit depends on:

- Sample size
- *RMSEA* cutoff values
  - Test of Close Fit: typically *RMSEA* = 0.05
  - Test of Not-Close Fit: typically *RMSEA* = 0.08 or 0.10
- Degrees of freedom in the model