Determining the number of factors using personality data

S. Mason Garrison
VANDERBILT
PEABODY COLLEGE

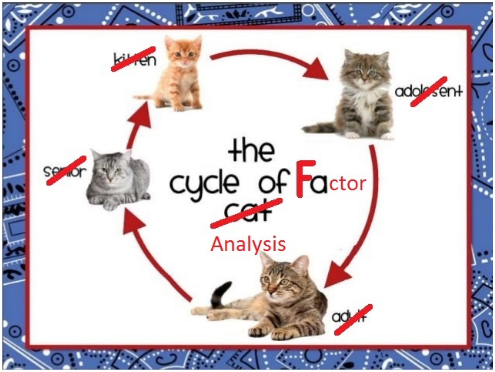
06/04/2017

Introduction

Differential Psychology and Factor Analysis

- The study of individual differences (*i.e.* intelligence, personality) is inexorably with factor analysis.
- Many generally accepted theories, such as the
    - General factor of Intelligence (Spearman, 1904) and the
    - Five Factor Model of Personality (**?**),
  - were developed, using factor analysis.
- Measures, based on those theories were developed,
  - using factor analysis.
- In turn, those measures were used to refine the theories,
  - which are used to create new measures...

- This cycle leaves Differential Psychology vulnerable



  - because factor analysis is misused often in construct validation research (Distefano & Hess, 2005),
  - conventional methods for determining the number of factors are subjective (Zwick & Velicer, 1986), and the
  - standard cut points for determining good fit aren't designed for personality-like data (Kang, McNeish, & Hancock, 2016).

Present Study

- The current study is designed to examine the effectiveness of factor enumeration rules on personality data;
- do the commonly used methods actually recover the correct number of factors?
- Specifically, how well do the following perform in recovering the correct number of factors?:
  - Minimum Average Partial procedure, and
  - various goodness-of-fit indices,
    - using classic (Hu & Bentler, 1998) thresholds.

Design Constants

- Five factors
- Estimated using MLE with Oblimin Rotation
- Adapted the vss function from Revelle's Psych package, using R 3.2.4 revised.
- Extracted maximum of 9 factors
- 100 Data Sets per condition

Data Generation

```
## Generate Data from factor loadings
# need a factor model and an effects matrix
GenData = function(fmodel,effect,n,names) {
  numberofvariables = dim(fmodel)[1]
  numberoflatent = dim(fmodel)[2]
  tmodel = t(fmodel)
  communality = diag(fmodel%*%tmodel)    #weight true
      scores and errors given the measurement model
  uniqueness = 1-communality
  errorweight = diag(sqrt(uniqueness))   #weight the
      errors
```

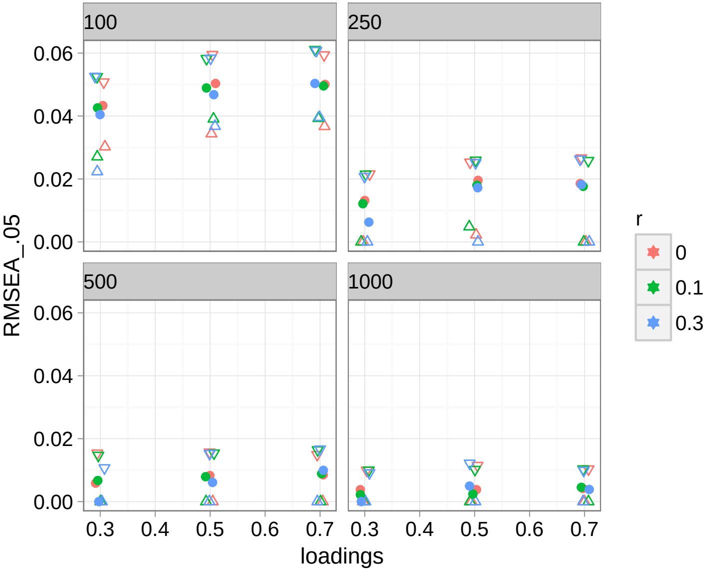
Data Generation

```
#create true scores for the latent variables
  latentscores = matrix(rnorm(n*(numberoflatent)),n)
  latentscores = latentscores%*%effect
  truescores = latentscores%*%tmodel

#create normal error scores
  error = matrix(rnorm(n*(numberofvariables)),n)
  error = error%*%errorweight

#create observed scores
  observedscore = truescores+error
  observedscore = data.frame(observedscore)
  names(observedscore)= names
  return(observedscore)}
```

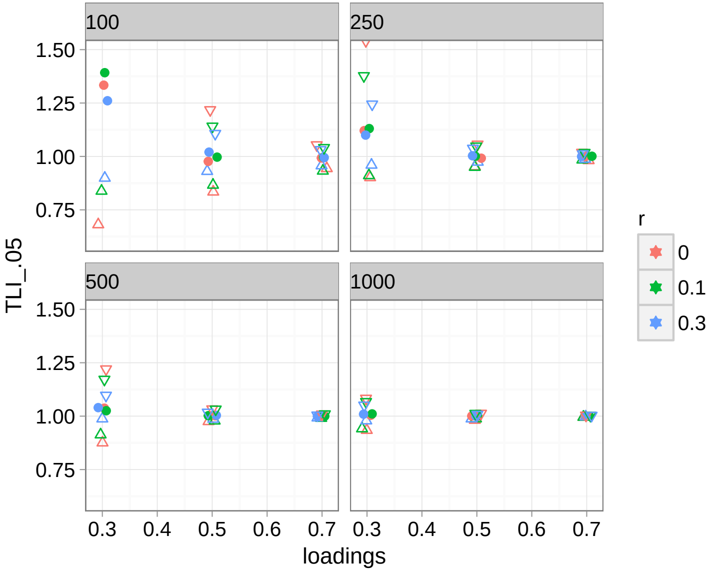
Design Conditions (3x3x3x4)

- Test Structure
  - Big Five Inventory (BFI; **?**)
    - 44 items, 16 reverse coded
  - International Personality Item Pool-NEO (IPIP-NEO; Goldberg, 1999)
    - 50 items, 24 reverse coded
    - 100 items, 47 reverse coded
- Item Loadings
  - .3, .5, .7
- Correlation between Factors
  - 0, .1, .3
- Sample Size
  - 100, 250, 500, 1000

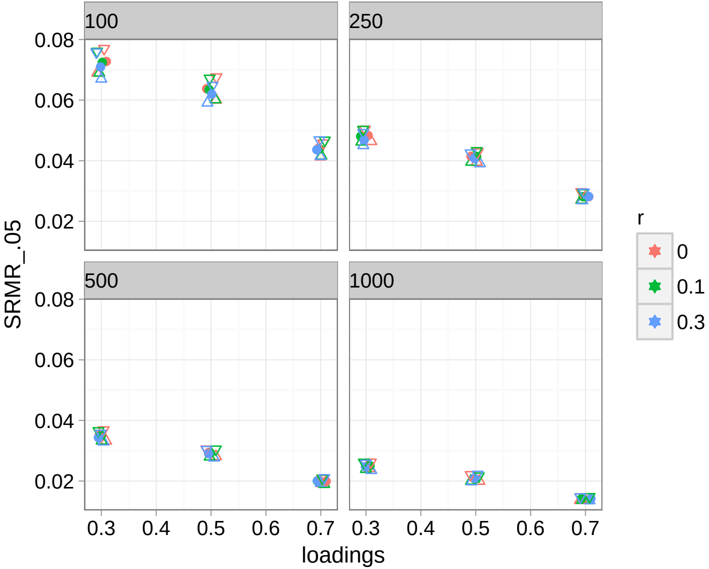
Fit Statistics for True Model:RMSEA



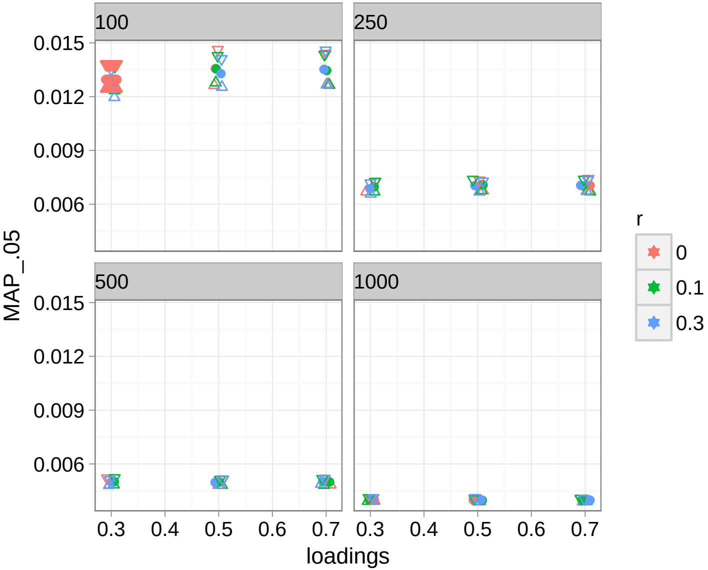
TLI


100
250
500
1000
1.50
1.25
1.00
0.75
TLI_.05
0.3
0.4
0.5
0.6
0.7
loadings
r
0
0.1
0.3

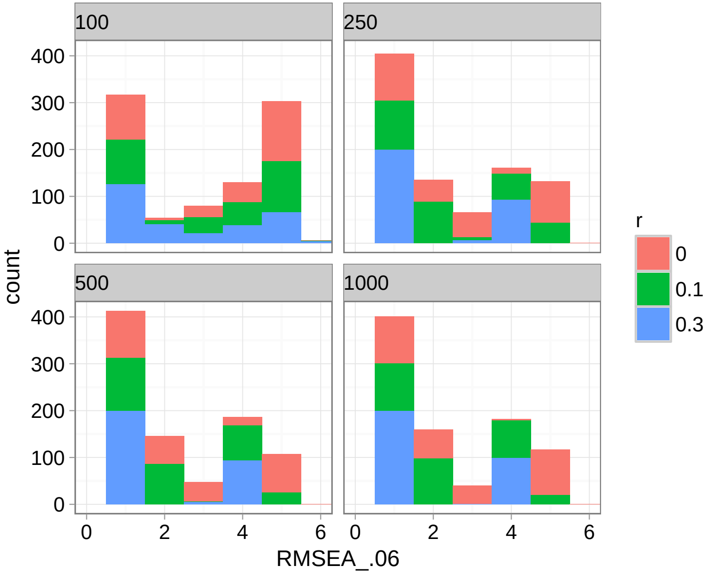
SRMR


100
250
500
1000
0.08
0.06
0.04
0.02
SRMR_.05
0.3
0.4
0.5
0.6
0.7
loadings
r
0
0.1
0.3

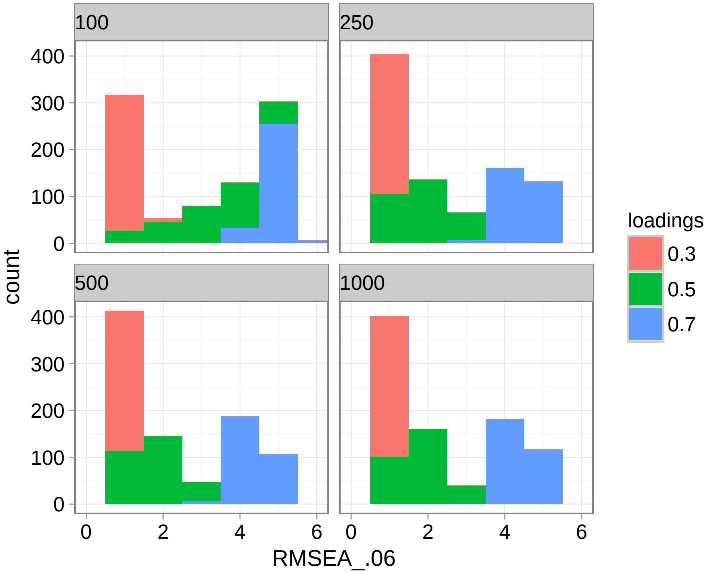
MAP


100
250
500
1000
0.015
0.012
0.009
0.006
MAP_.05
0.3
0.4
0.5
0.6
0.7
loadings
r
0
0.1
0.3

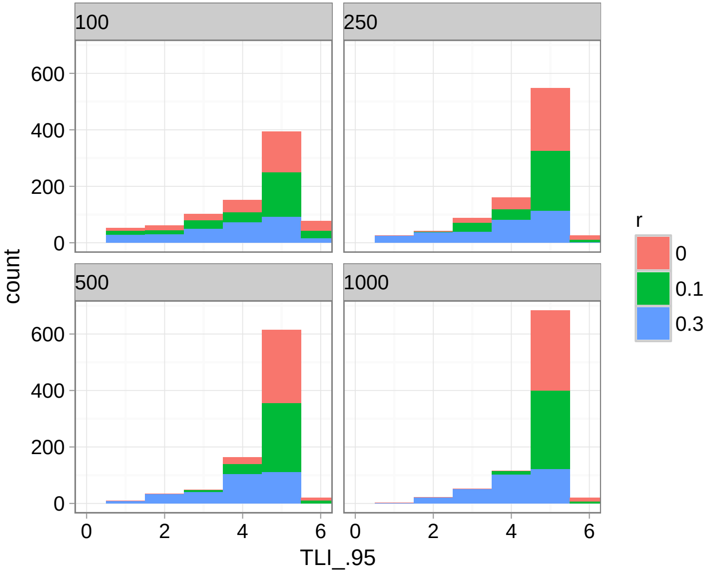
RMSEA Point Estimate by Correlated Factors



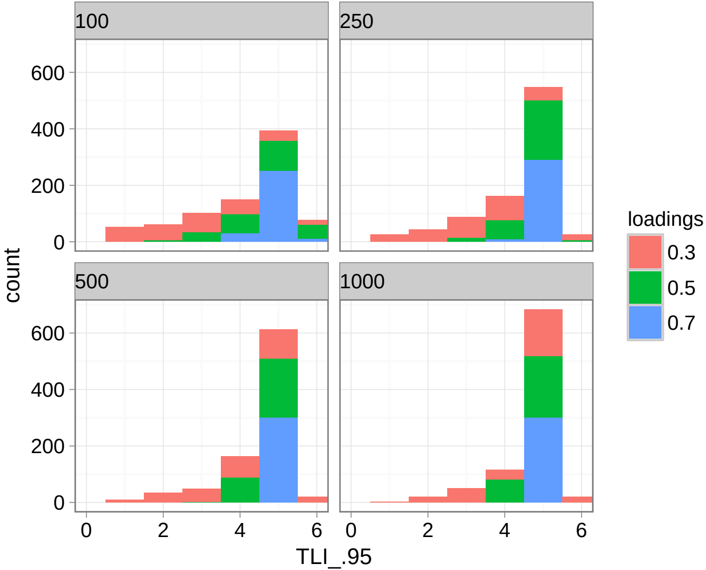
RMSEA Point by Loading



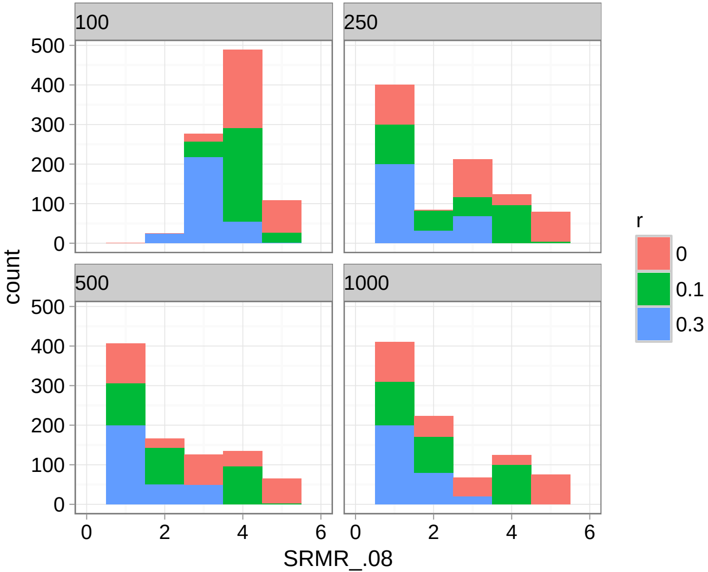
TLI by Correlated Factors



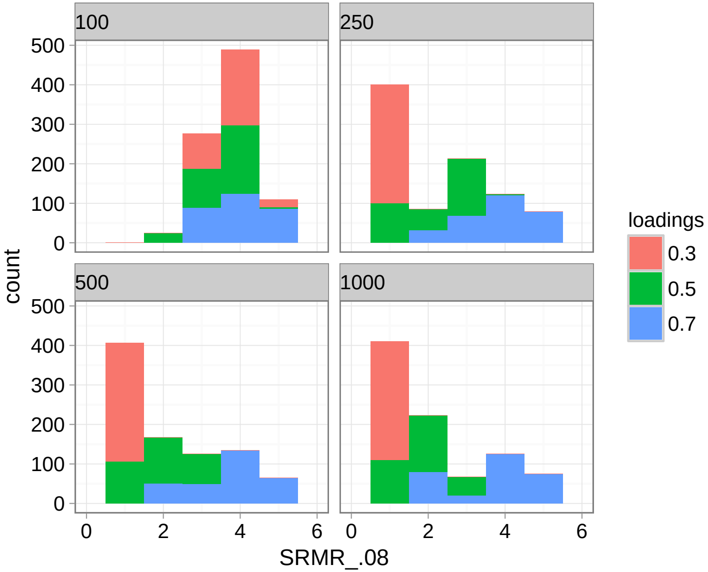
TLI by Loading



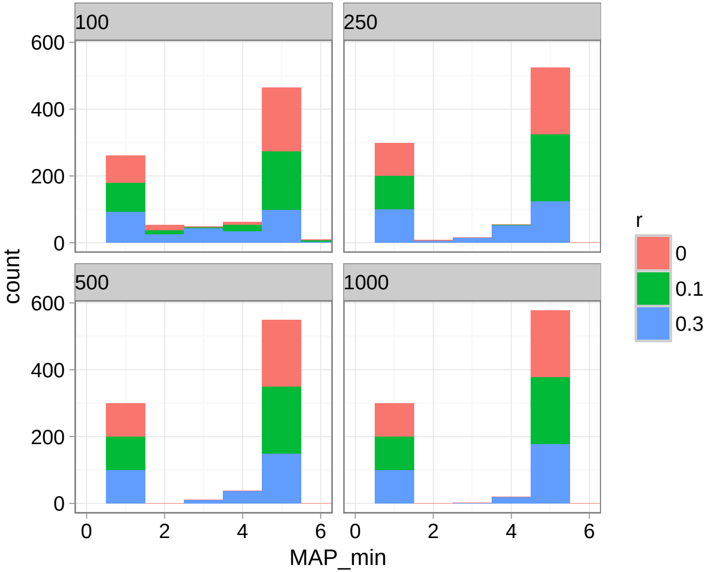
SRMR by Correlated Factors



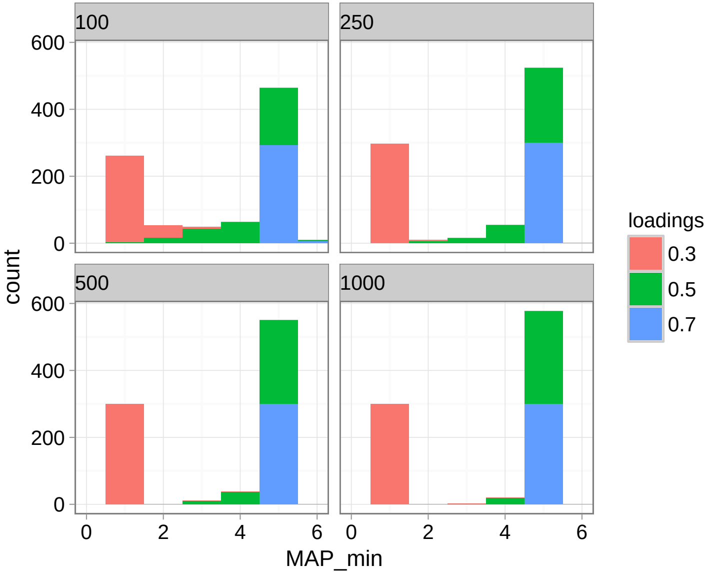
SRMR by Loading



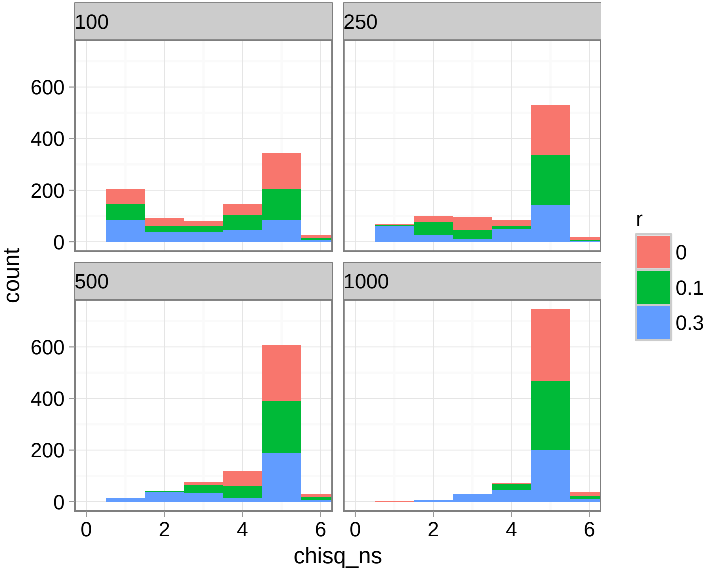
MAP by Correlated Factors



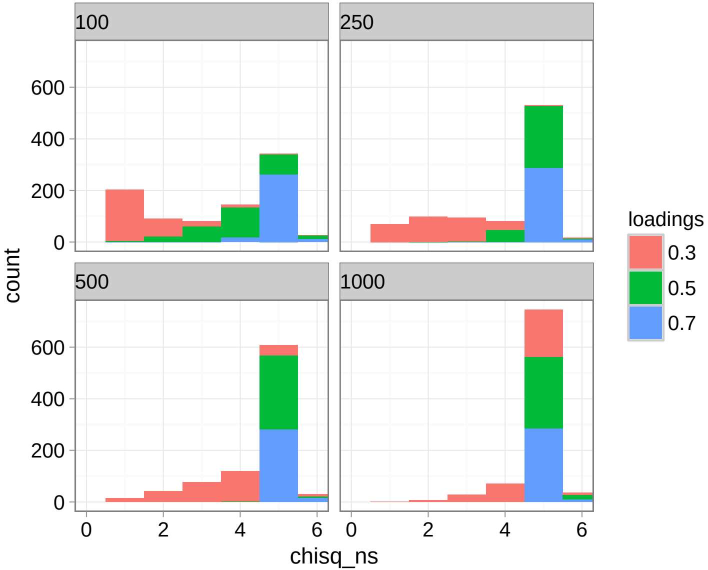
MAP by Loading



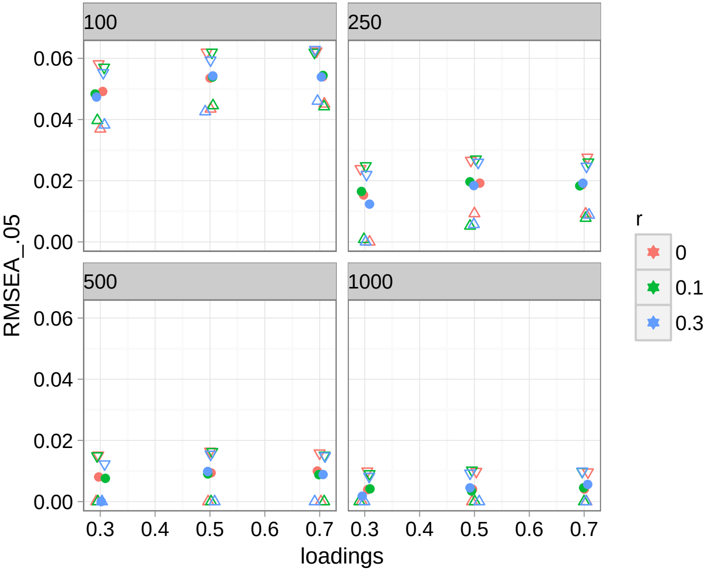
CHI^2 BY CORRELATED FACTORS



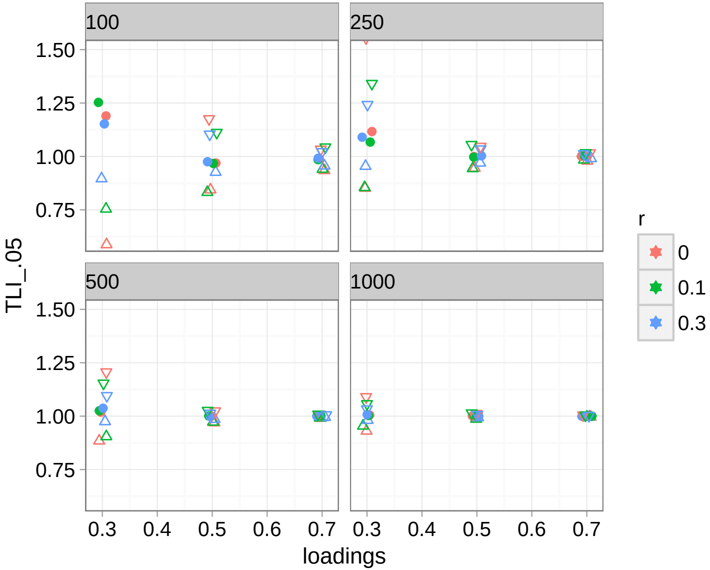
Chi^2 by Loading



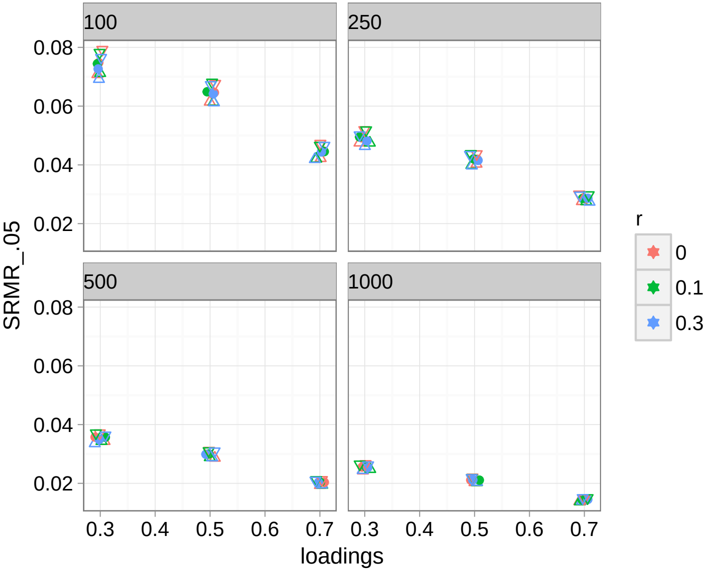
Fit Statistics for True Model:RMSEA



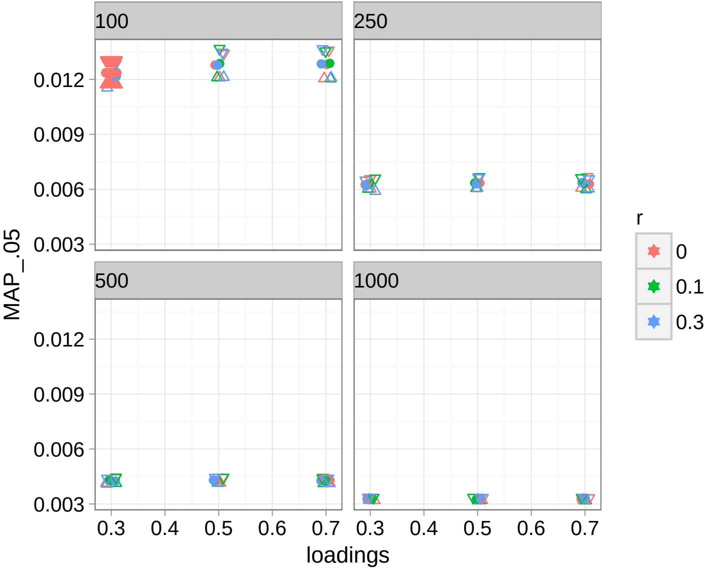
TLI


100
250
500
1000
1.50
1.25
1.00
0.75
TLI_.05
0.3
0.4
0.5
0.6
0.7
loadings
r
0
0.1
0.3

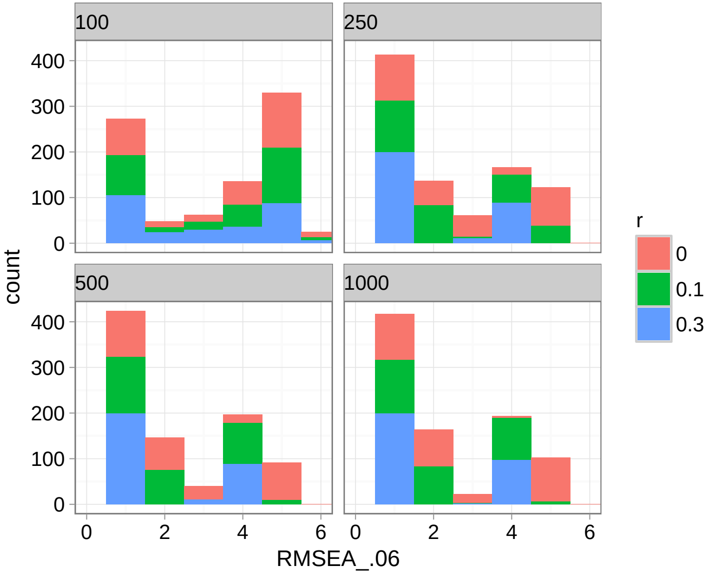
SRMR


100
250
500
1000
0.08
0.06
0.04
0.02
SRMR_.05
0.3
0.4
0.5
0.6
0.7
loadings
r
0
0.1
0.3

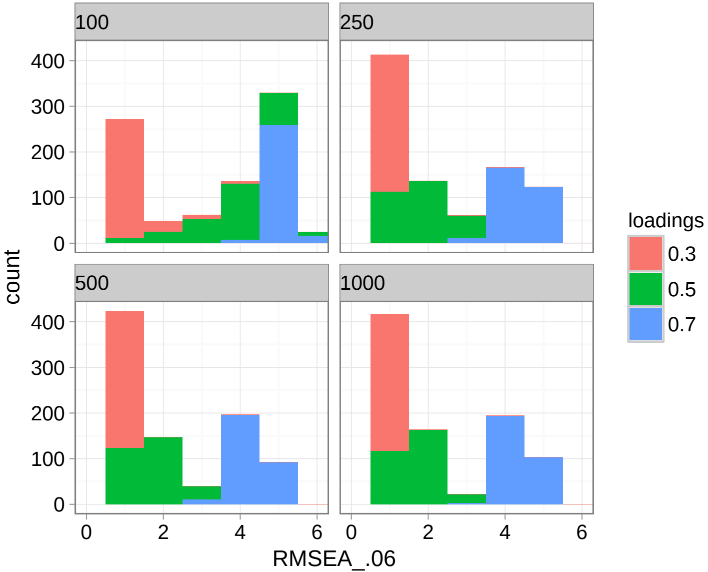
MAP


100
250
500
1000
MAP_.05
0.003
0.006
0.009
0.012
loadings
0.3
0.4
0.5
0.6
0.7
r
0
0.1
0.3

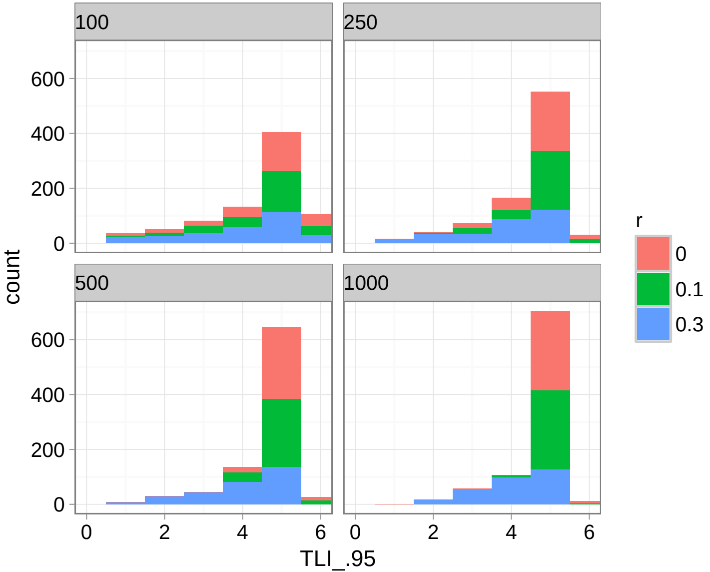
RMSEA Point Estimate by Correlated Factors



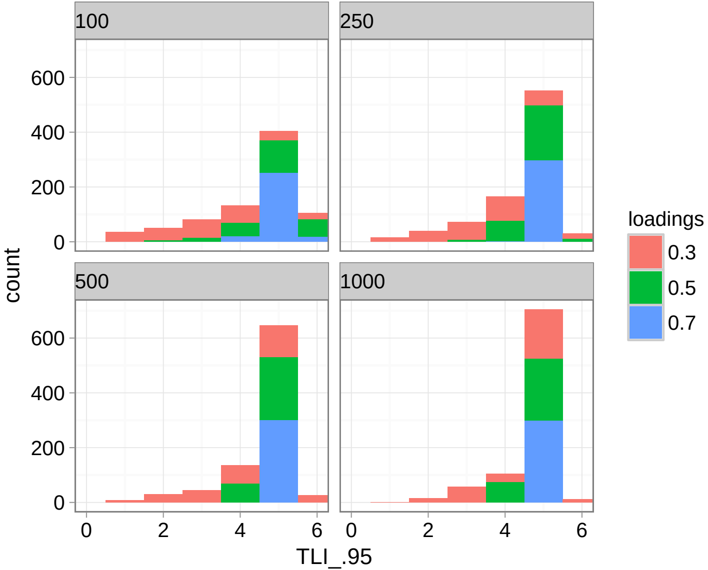
RMSEA Point by Loading



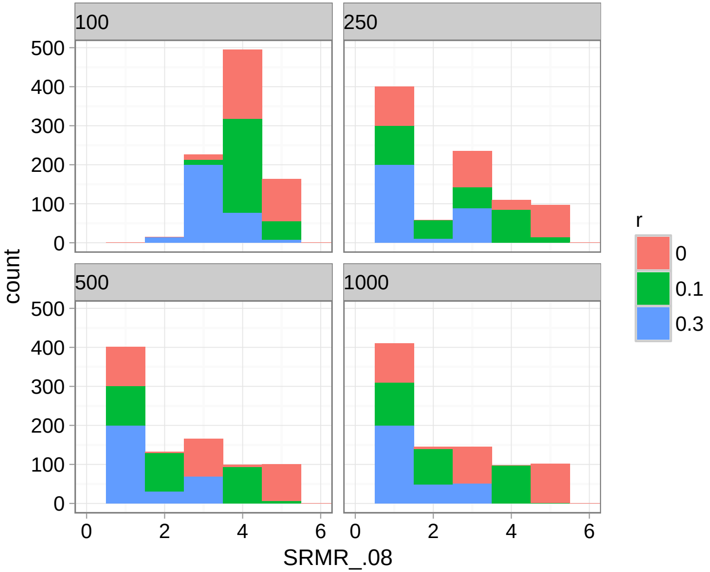
TLI by Correlated Factors



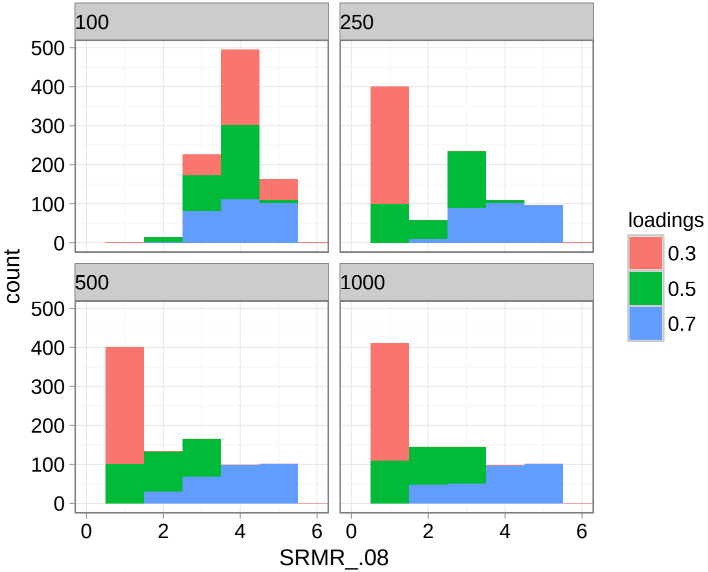
TLI by Loading



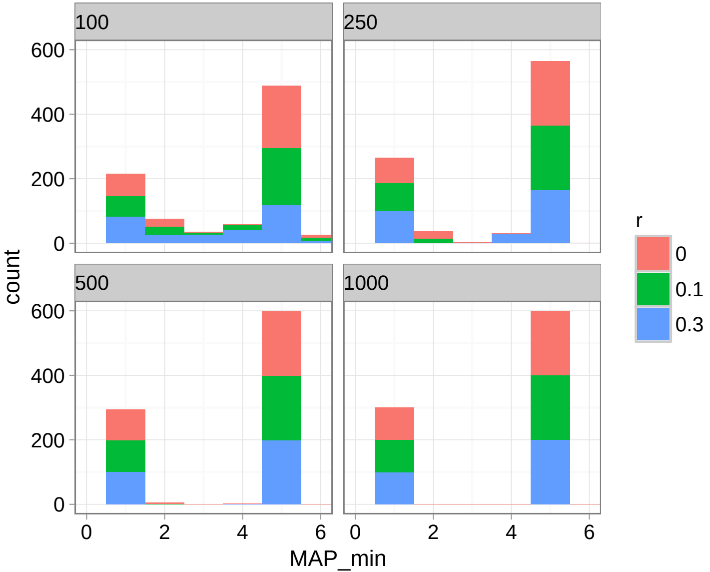
SRMR by Correlated Factors



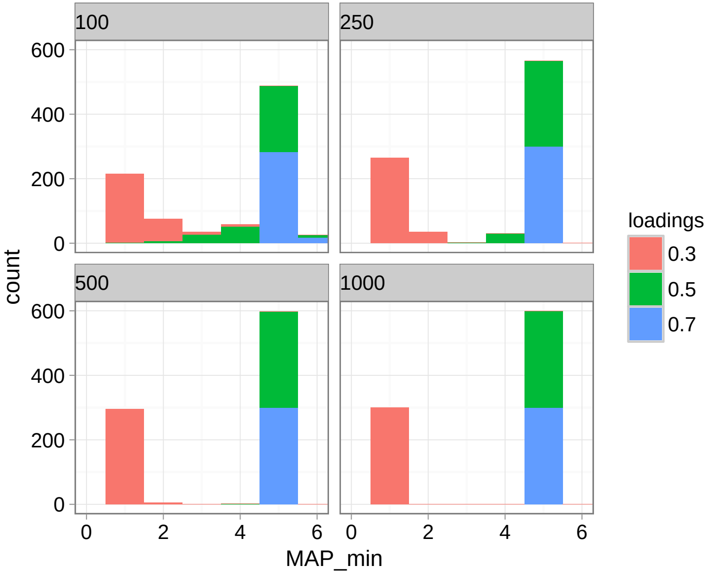
SRMR by Loading



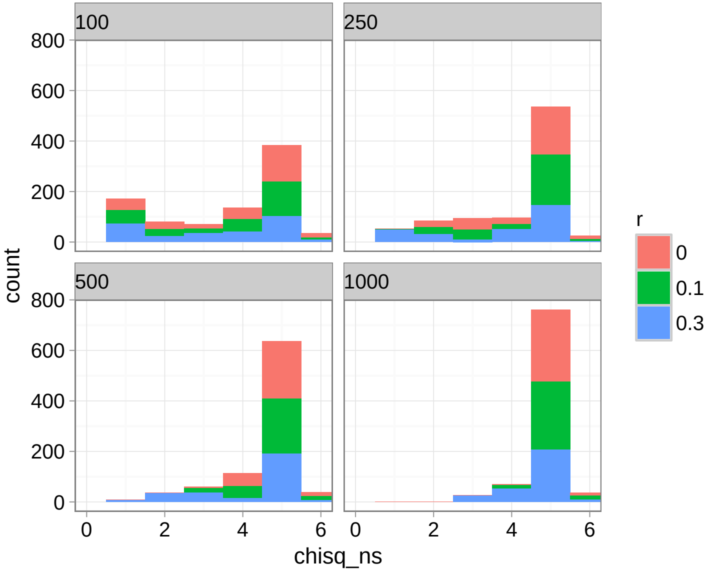
MAP by Correlated Factors



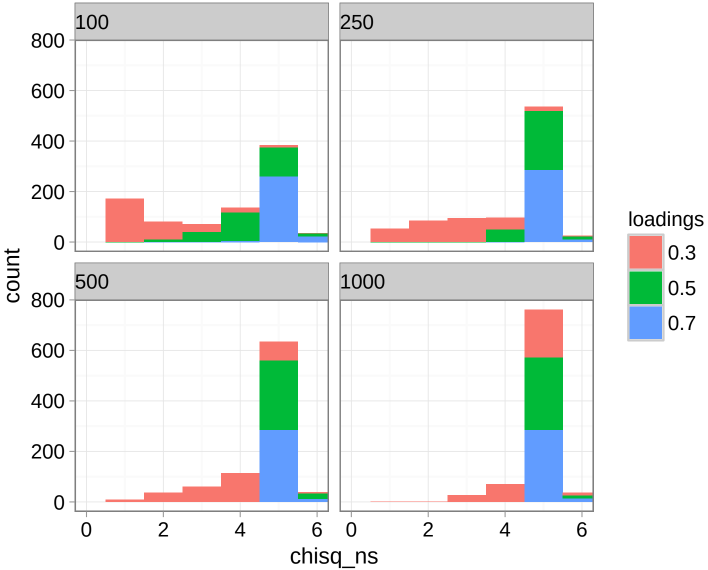
MAP by Loading



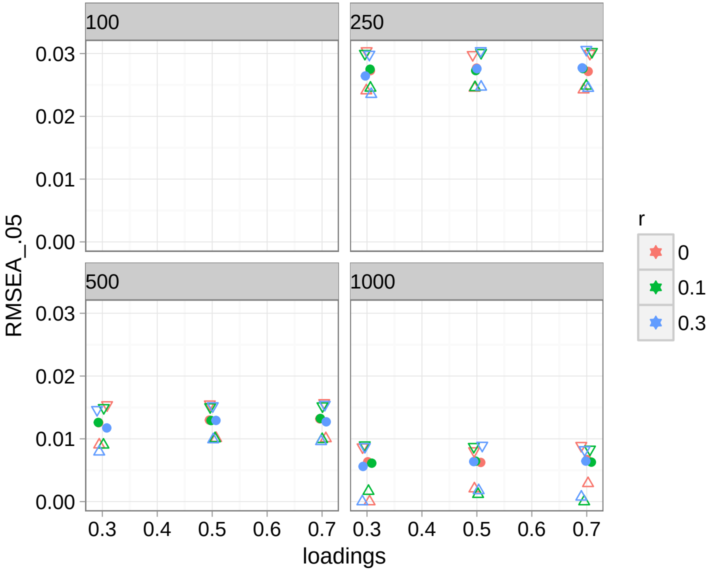
Chi^2 by Correlated Factors



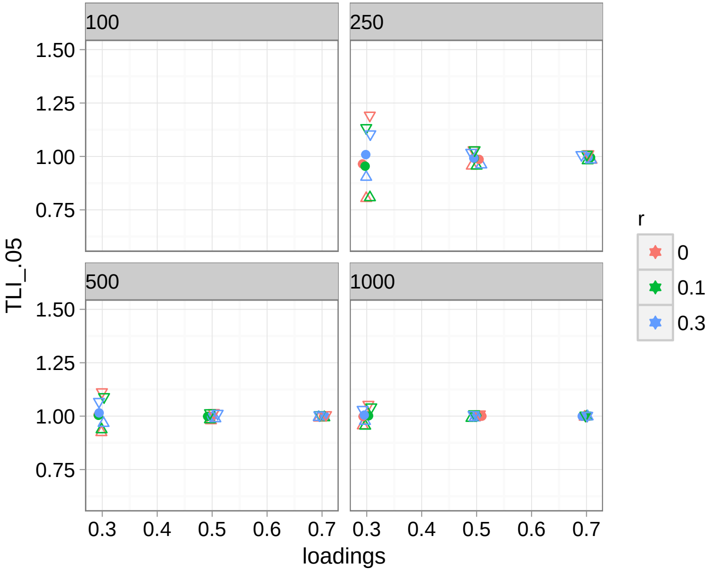
Chi^2 by Loading



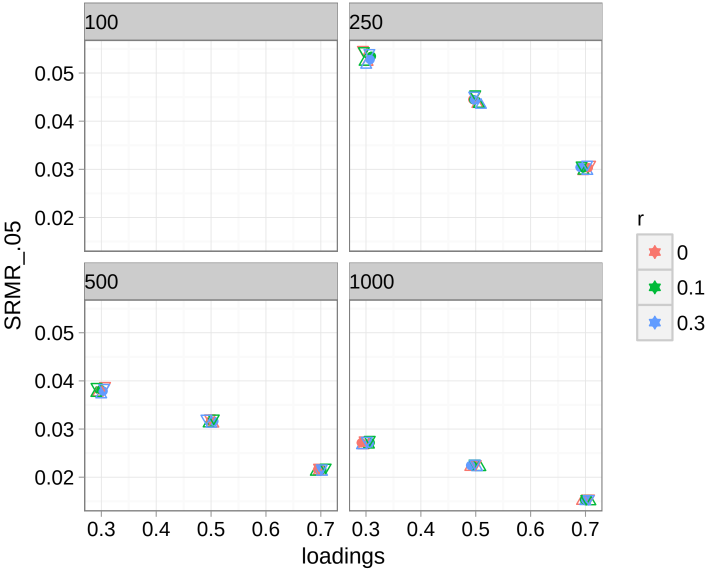
Fit Statistics for True Model:RMSEA



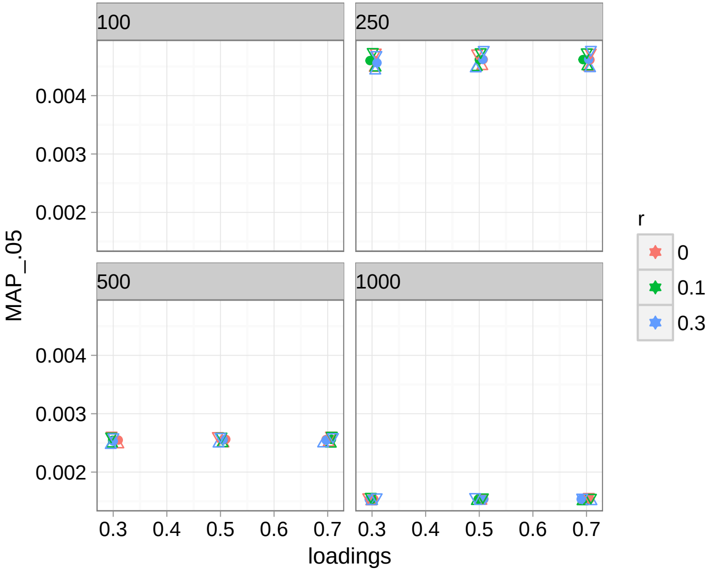
TLI


100
250
500
1000
1.50
1.25
1.00
0.75
TLI_.05
0.3
0.4
0.5
0.6
0.7
loadings
r
0
0.1
0.3

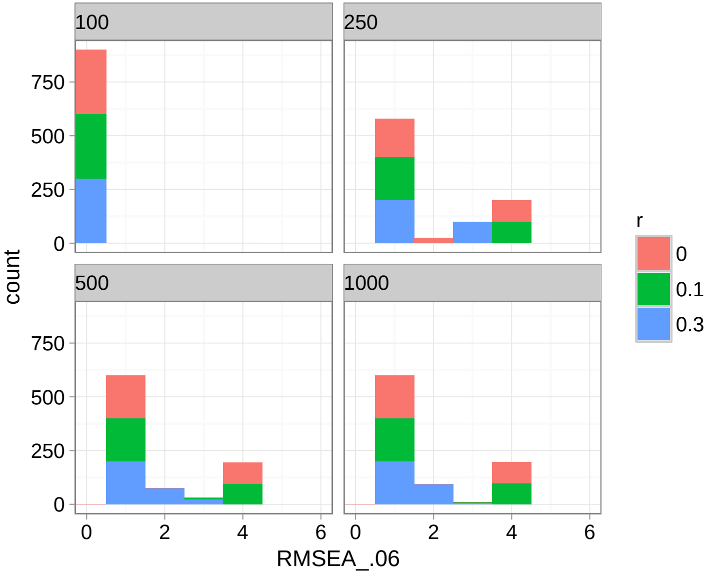
SRMR


100
250
500
1000
0.05
0.04
0.03
0.02
SRMR_.05
0.3
0.4
0.5
0.6
0.7
loadings
r
0
0.1
0.3

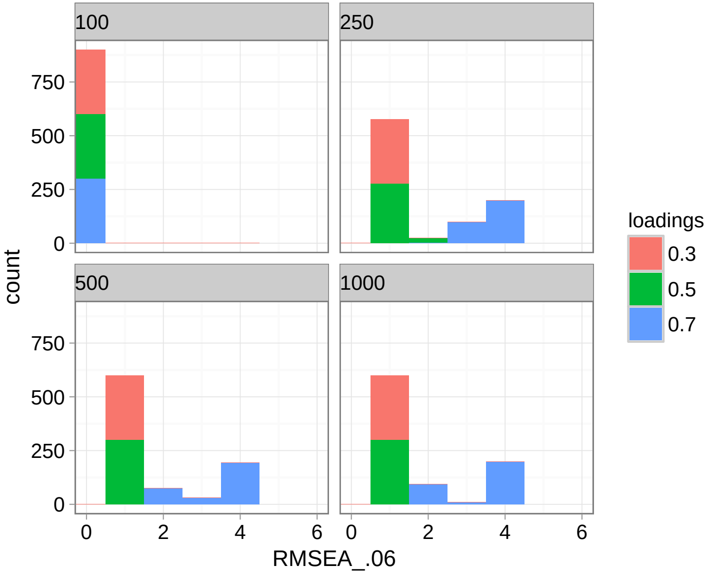
MAP


100
250
500
1000
0.004
0.003
0.002
MAP_.05
0.3
0.4
0.5
0.6
0.7
loadings
r
0
0.1
0.3

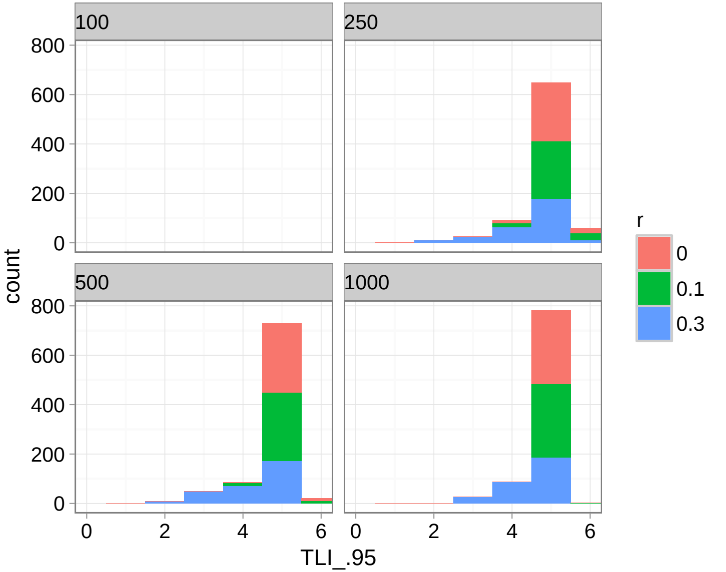
RMSEA Point Estimate by Correlated Factors



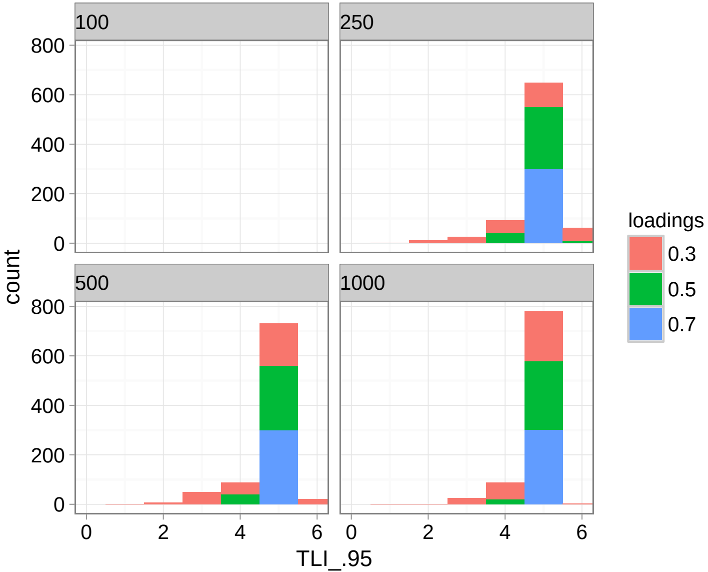
RMSEA Point by Loading



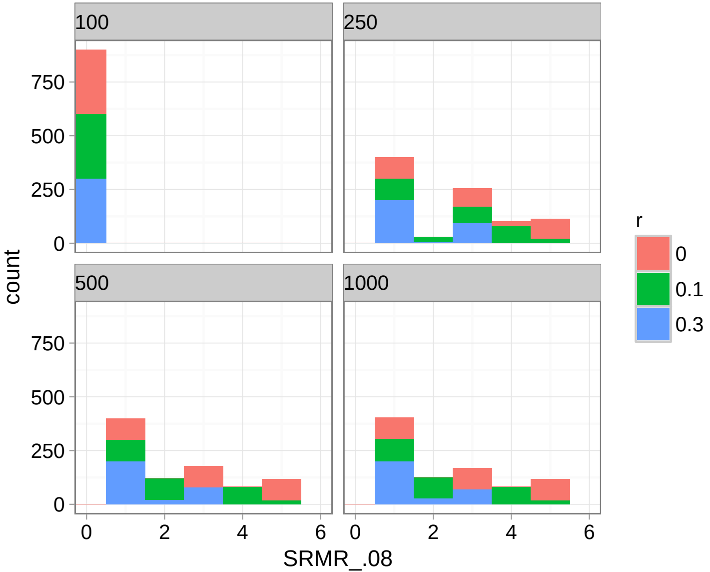
TLI by Correlated Factors



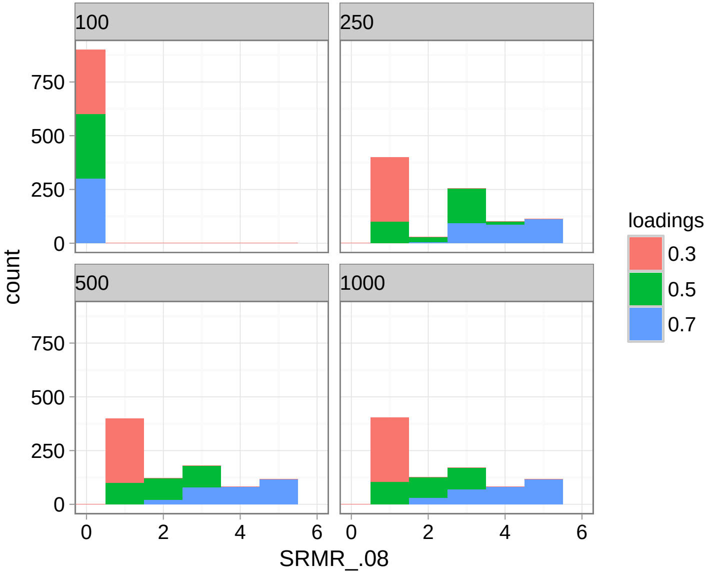
TLI by Loading



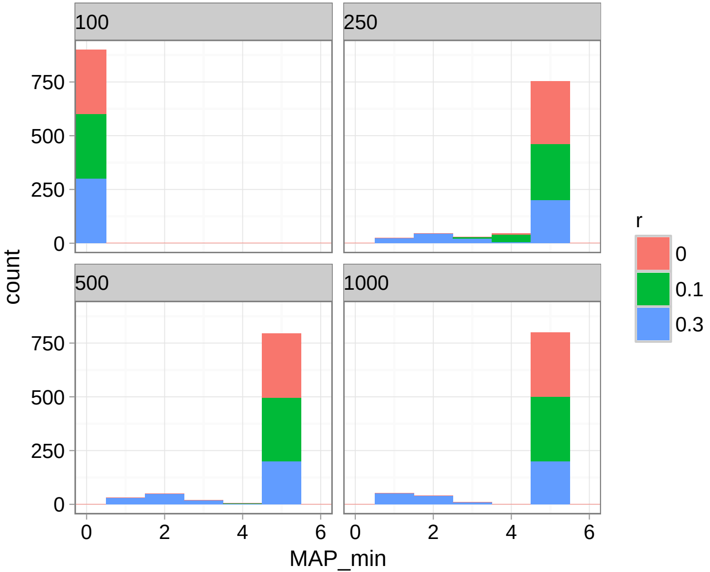
SRMR by Correlated Factors



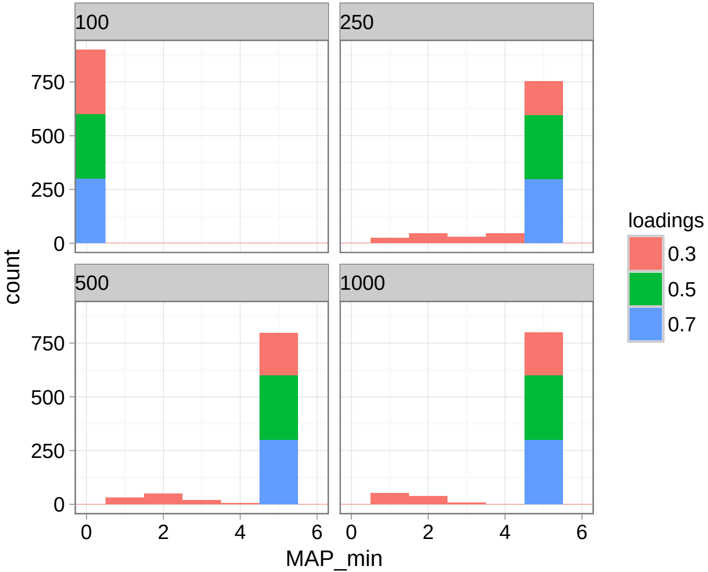
SRMR BY LOADING



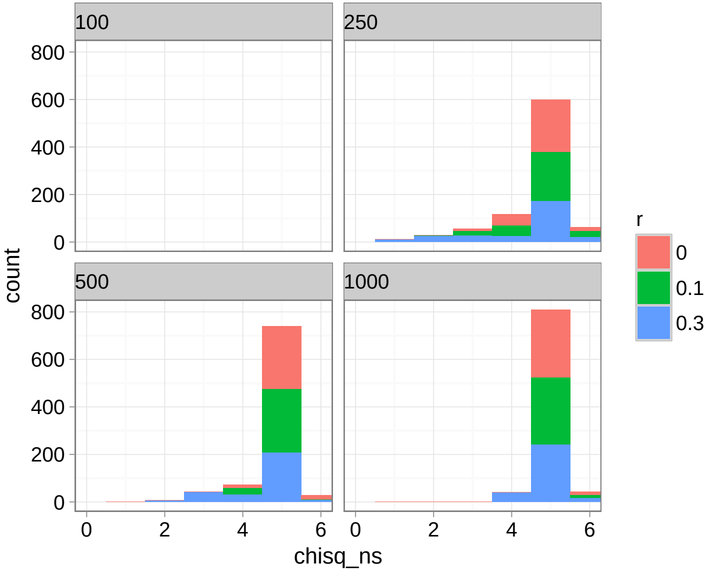
MAP by Correlated Factors



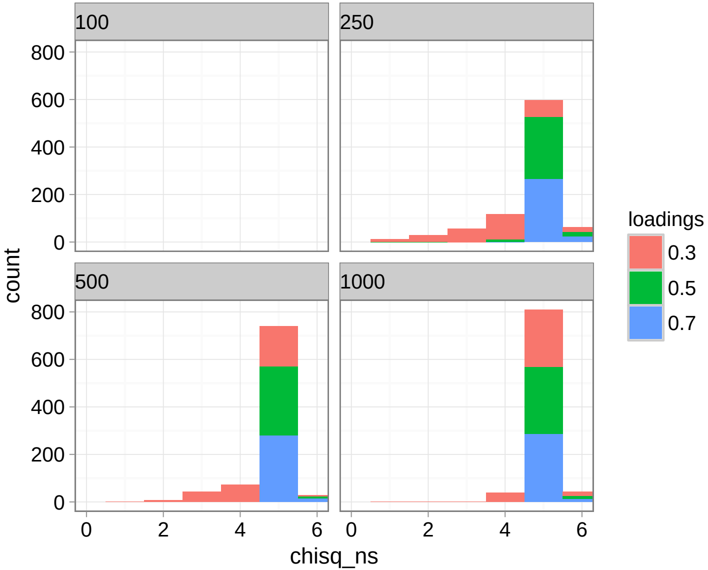
MAP by Loading



CHI^2 BY CORRELATED FACTORS



CHI^2 BY LOADING



Discussion

- Interpretation
  - Tendency to underfactor
  - High Correlations between factors
  - Low Factor Loadings
- Recommendations
  - Sample Size
  - Measurement Matters
- Future Directions

References

Distefano, C., & Hess, B. (2005). Using confirmatory factor analysis for construct validation : An empirical review. *Journal of Psychoeducational Assessment*, 225–241.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality psychology in Europe*, *7*, 7–28. Retrieved from `http://projects.ori.org/lrg/PDFs_papers/Abroad-bandwidthinventory.pdf`

Hu, L.-t., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424–453.

Kang, Y., McNeish, D. M., & Hancock, G. R. (2016). The Role of Measurement Quality on Practical Guidelines for Assessing Measurement and Structural Invariance. *Educational and Psychological Measurement*.

Spearman, C. (1904). " General Intelligence," objectively determined and measured. *The American Journal of Psychology*, *15*(2), 201–292.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*(3), 432.