UNIVERSITY OF ARKANSAS

Regularization for Variable Selection in Multidimensional Factor and Network Models

Jiaying Chen
Xinya Liang, Ph.D
Jihong Zhang, Ph.D

Introduction

- Variable selection is crucial for achieving model simplicity by reducing the number of items included in the analysis, helping to create a more interpretable and efficient model (Jacobucci et al., 2016 ;Wille, 1996).

- Both FA and Network Analysis can incorporate regularization techniques (e.g. Lasso) to aid in simplifying models by penalizing less important parameters, thereby enhancing model estimation and preventing potential overfitting.

- How regularized factor and network analysis perform in terms of variable selection in instruments developed through the factor analytic framework, particularly across unidimensional and multidimensional structures, remains understudied.

Background information

Regularized Factor Analysis Models:

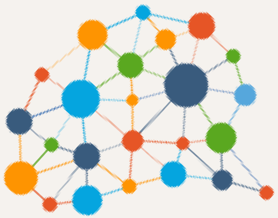
Factor analysis models seek to uncover latent structures that explain relationships among observed variables. Regularization is a technique integrated into the estimation process to enhance the model by including a penalty term with the maximum likelihood estimation objective function.

$$F_{FA} = F_{ML} + \gamma * P(\cdot),$$

Various penalty terms P() are available to achieve regularization.

- $\gamma P_{lasso}(\theta) = \gamma \, \|\theta\|_1 = \gamma \sum|\theta|$ (Tibshirani, 1996)
- $\gamma P_{Enet}(\theta) = \gamma * [(1-\alpha)\|\theta\|_1 + \alpha * \|\theta\|_2]$ (Zou & Hastie, 2005)

Background information



Regularized Psychometric Network Models:

Psychological networks consist of nodes representing observed variables, connected by edges representing statistical relationships (Epskamp et al., 2017) .

- Pruning: Shrink less important connections to zero, which helps in focusing on the most influential relationships. (Han et al., 2015)

- $F_{Glasso}(\Theta) = -\log det\,(\Theta) + trace(S\Theta) + \lambda \sum_{i\neq j}\left|\Theta_{ij}\right|$ (Friedman et al., 2008)

Purpose of study

- **Aim 1:** Compare the efficacy of variable selection in regularized factor analysis vs. network analysis under single and multidimensional contexts.

- **Aim 2:** Explore how regularized network models can inform variable selection in data generated with latent structures.

Study design

Data were generated in R 4.3.3 with the *lavaan* package (Rosseel, 2012).

- Population model:

  one-factor model with 24 normally distributed items (0.7 /0.2, 0.1, 0) .

  three-factor model with 8 items for each factor (0.7 /0.2, 0.1, 0).

  three-factor model with 8 items for each factor (one cross loading each factor)

- Sample size : 200, 500, 1000

- Replications: 100

Data analysis

Data were analyzed in R 4.3.3 with the following packages (Rosseel, 2012) .

- Network Analysis Method:

  Prune in '**psychonetrics**': τ =0.01

  Glasso in '**bootnet**': γ= 0.5

- Factor Analysis Method:

  Lasso and Elastic net in 'regsem', 'penfa', and 'lslx' packages

  lambda.start= 0.001, jump = 0.01, n.lambda = 50

  Elastic net alpha = 0.5

  Threshold = 0.1 (zhang, 2021)

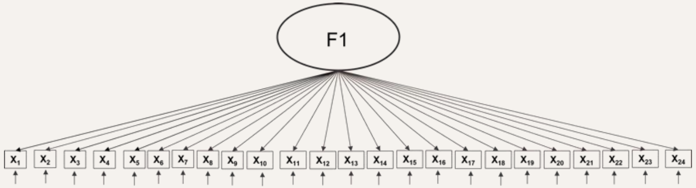
R package

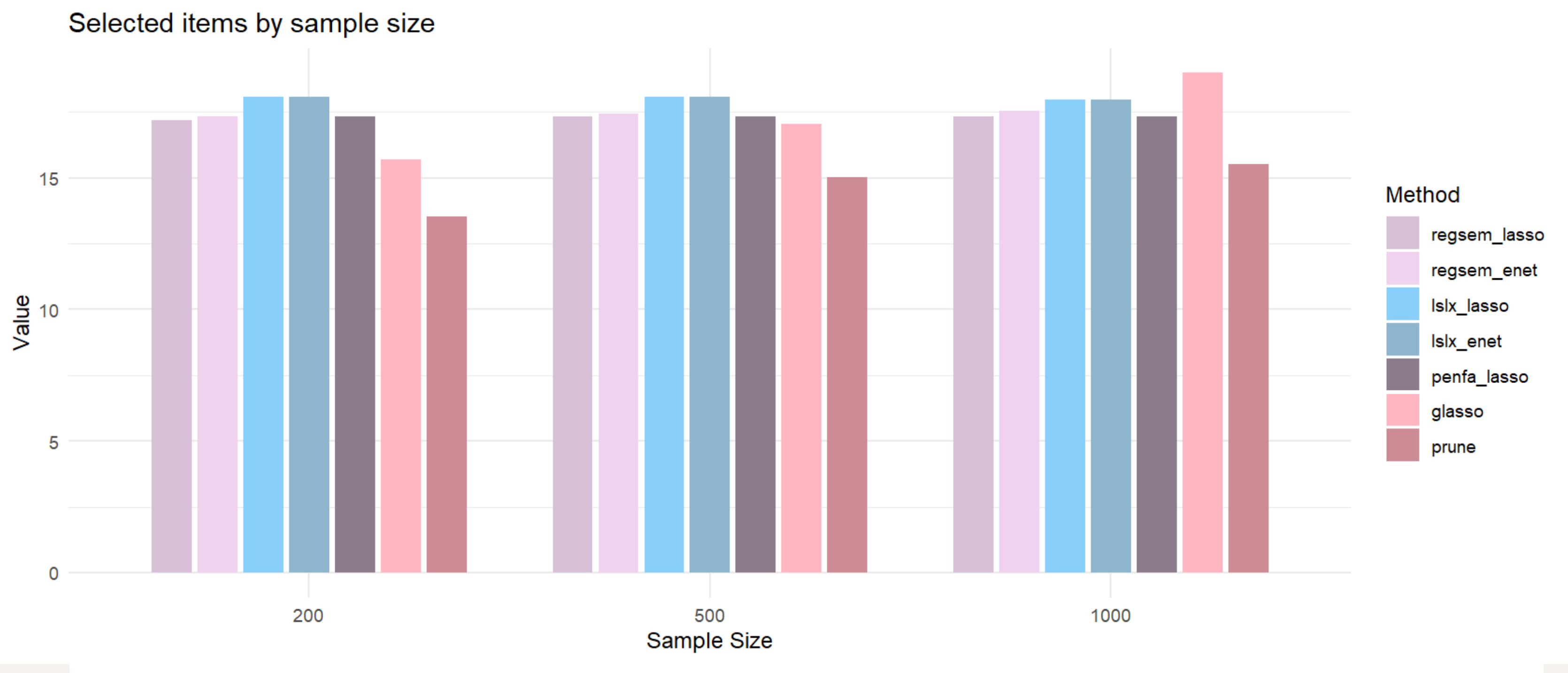
	Factor analysis	
	Lasso	ElasticNet
regsem	√	√
panfa	√	--
Lslx	√	√

	Network analysis
psychonetrics	prune
bootnet	glasso

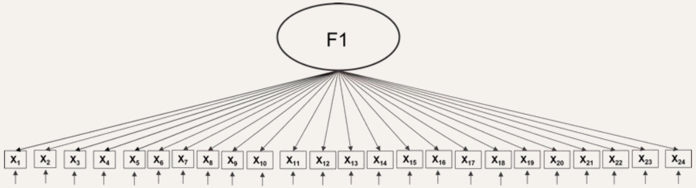
Results (one factor model)

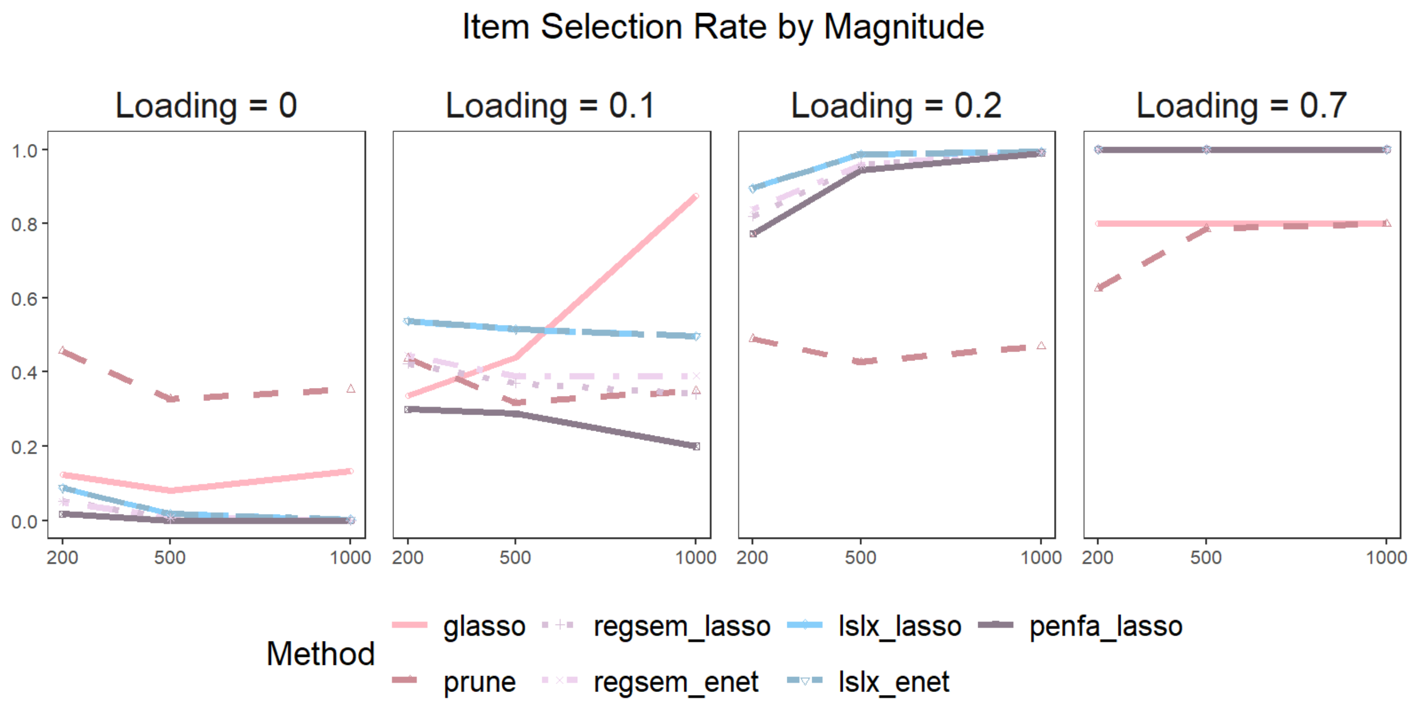
Non-zero items = 20





Results (one factor model)





$$\text{TP rate} = \frac{TP}{TP+FN}$$

$$\text{FP rate} = \frac{FP}{FP+TN}$$

Results (one factor model)

BIC 200

	regsem	penfa	lslx
lasso	12647.57	12671.26	-5.32
enet	12648.74	--	-5.32
	psycho	**bootnet**	
prune	13359.13	--	
glasso	--	4530.318	

BIC 500

	regsem	penfa	lslx
lasso	31389.28	31415.94	-2.60
enet	31390.26	--	-2.61
	psycho	**bootnet**	
prune	32219.32	--	
glasso	--	10235.49	

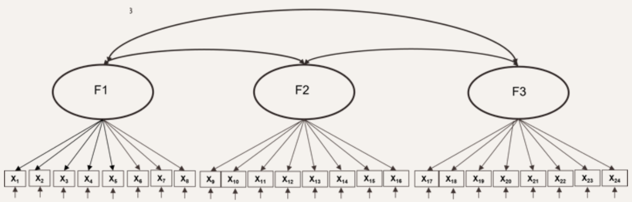
BIC 1000

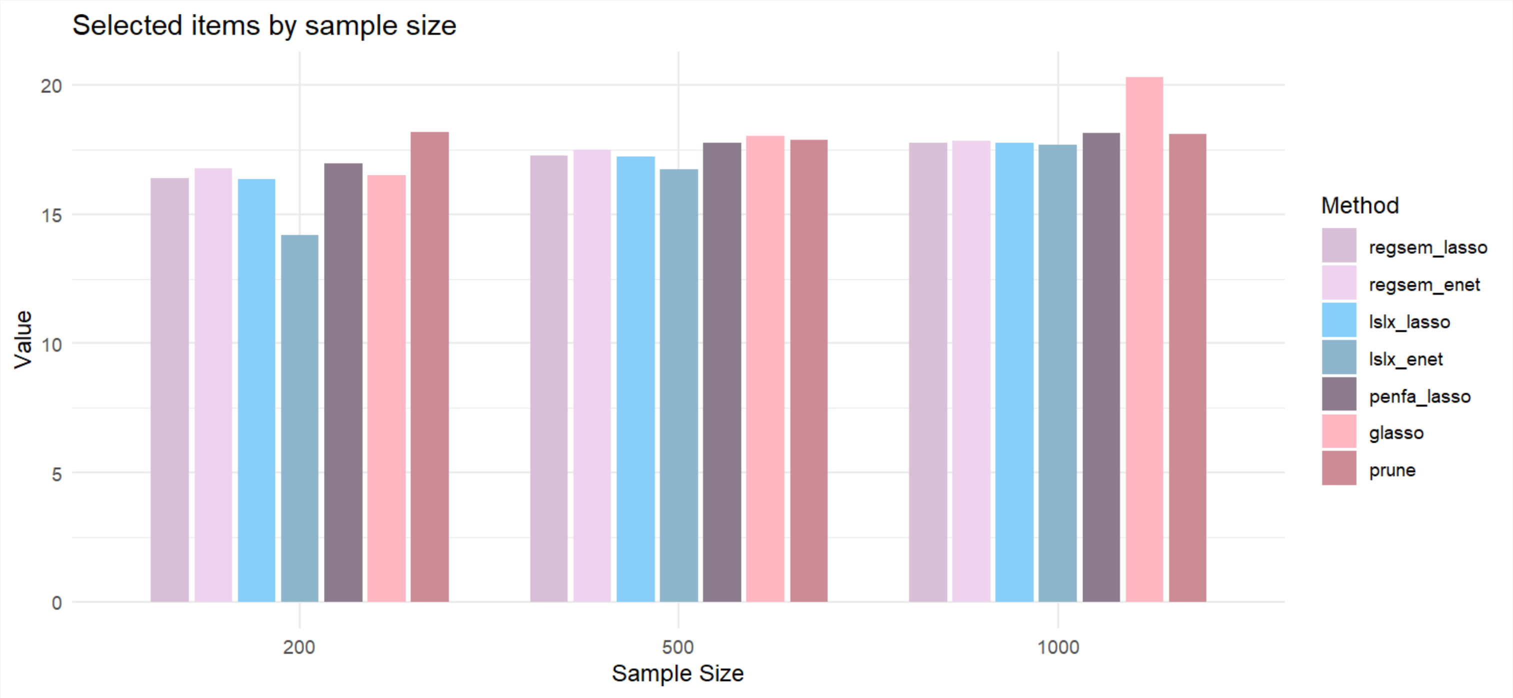
	regsem	penfa	lslx
lasso	62637.7	62642.92	-1.48
enet	62638.63	--	-1.49
	psycho	**bootnet**	
prune	63230.48	--	
glasso	--	19606.96	

For Glasso, the number demonstrate EBIC

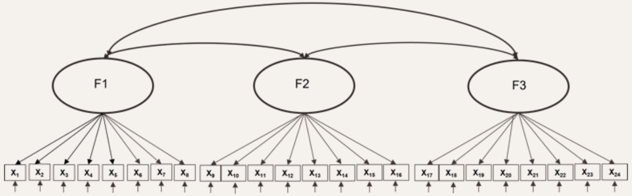
Results (three factor model)

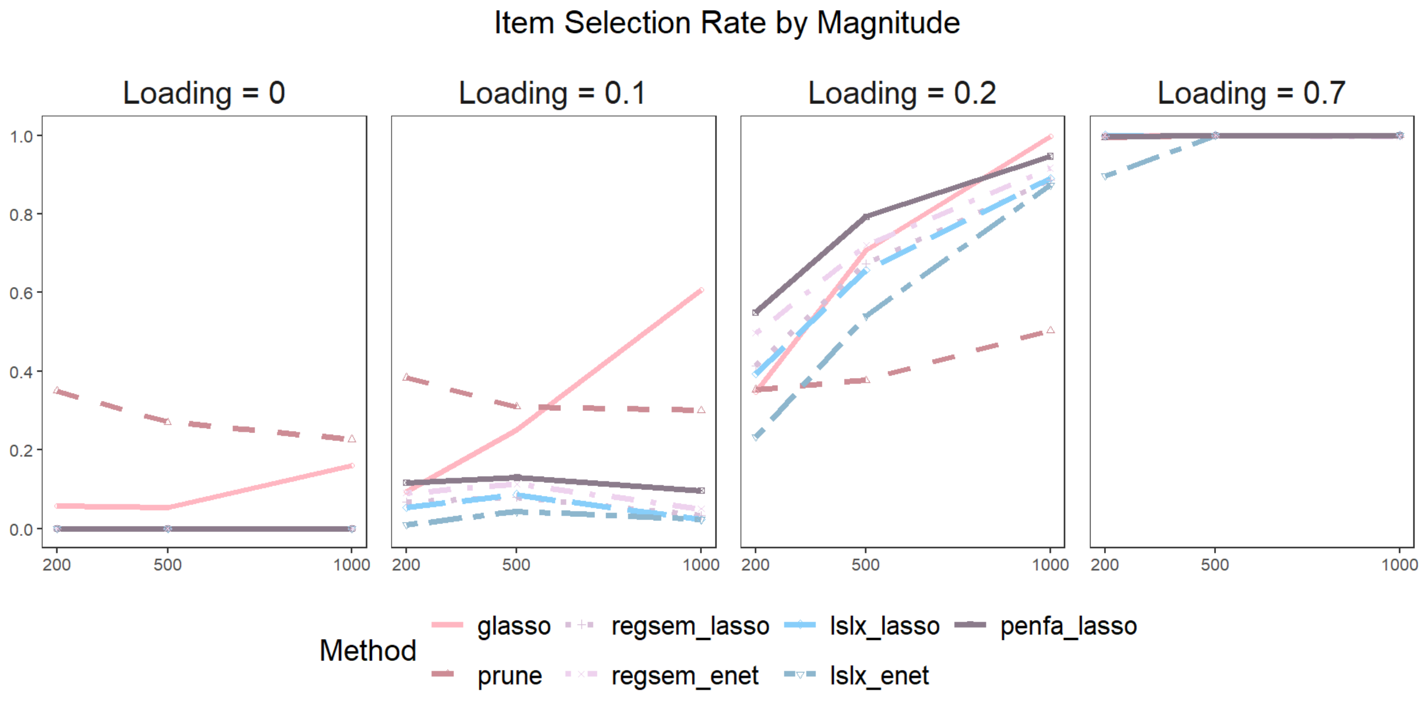
Non-zero items = 21





Results (three factor model)





TP rate = $\frac{TP}{TP+FN}$

FP rate = $\frac{FP}{FP+TN}$

Results (three factor model)

BIC 200

	regsem	penfa	lslx
lasso	12839.57	12809.32	-5.00
enet	12856.84	---	-5.31
	psycho	bootnet	
prune	13008.36	--	
glasso	--	4404.95	

BIC 500

	regsem	penfa	lslx
lasso	31795.53	31754.46	-2.44
enet	31813.92	---	-2.49
	psycho	bootnet	
prune	32010.30	--	
Glasso	--	10283.80	

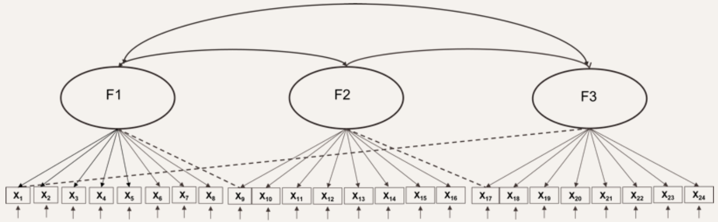
BIC 1000

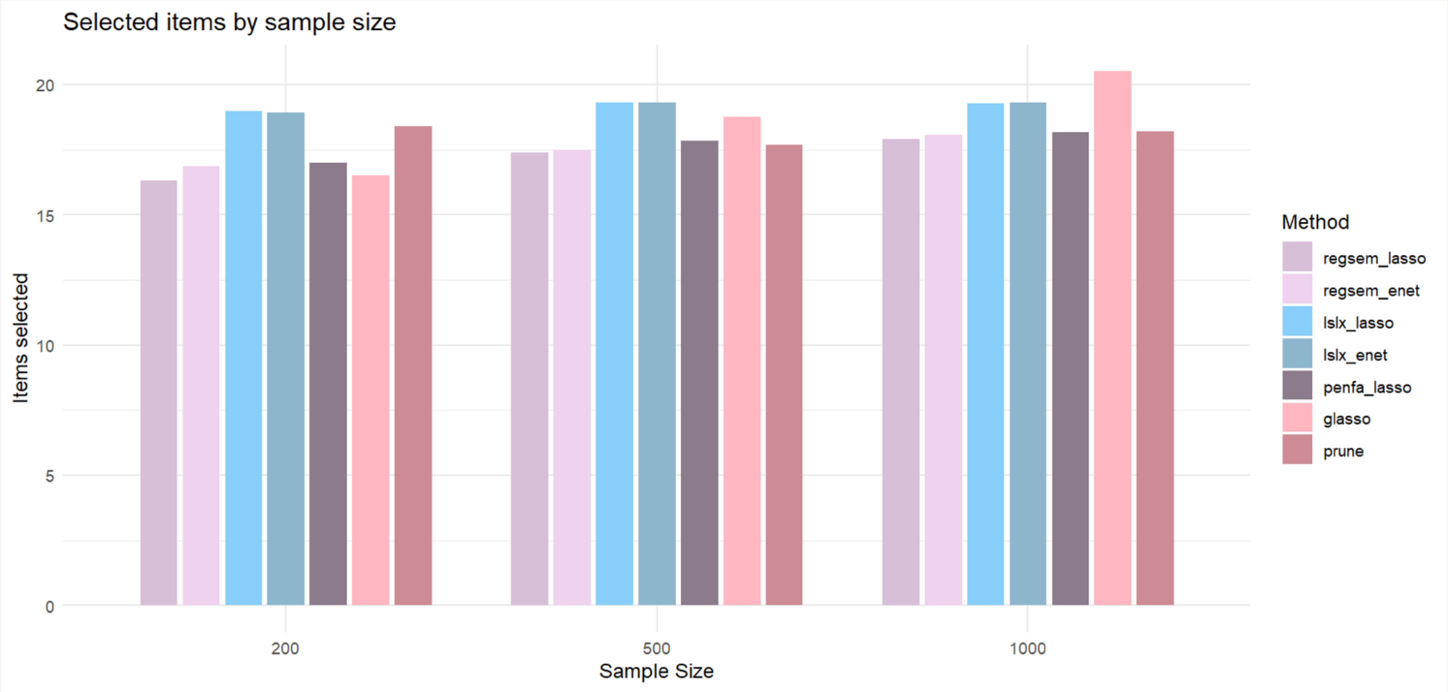
	regsem	penfa	lslx
lasso	63373.67	63345.49	-1.38
enet	59539.67	---	-1.38
	psycho	bootnet	
prune	63700.98	--	
glasso	--	20015.60	

For Glasso, the number demonstrate EBIC

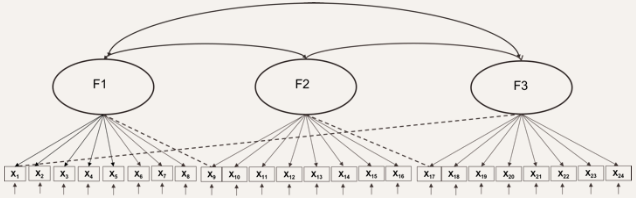
Results (three factor model with cross-loading)

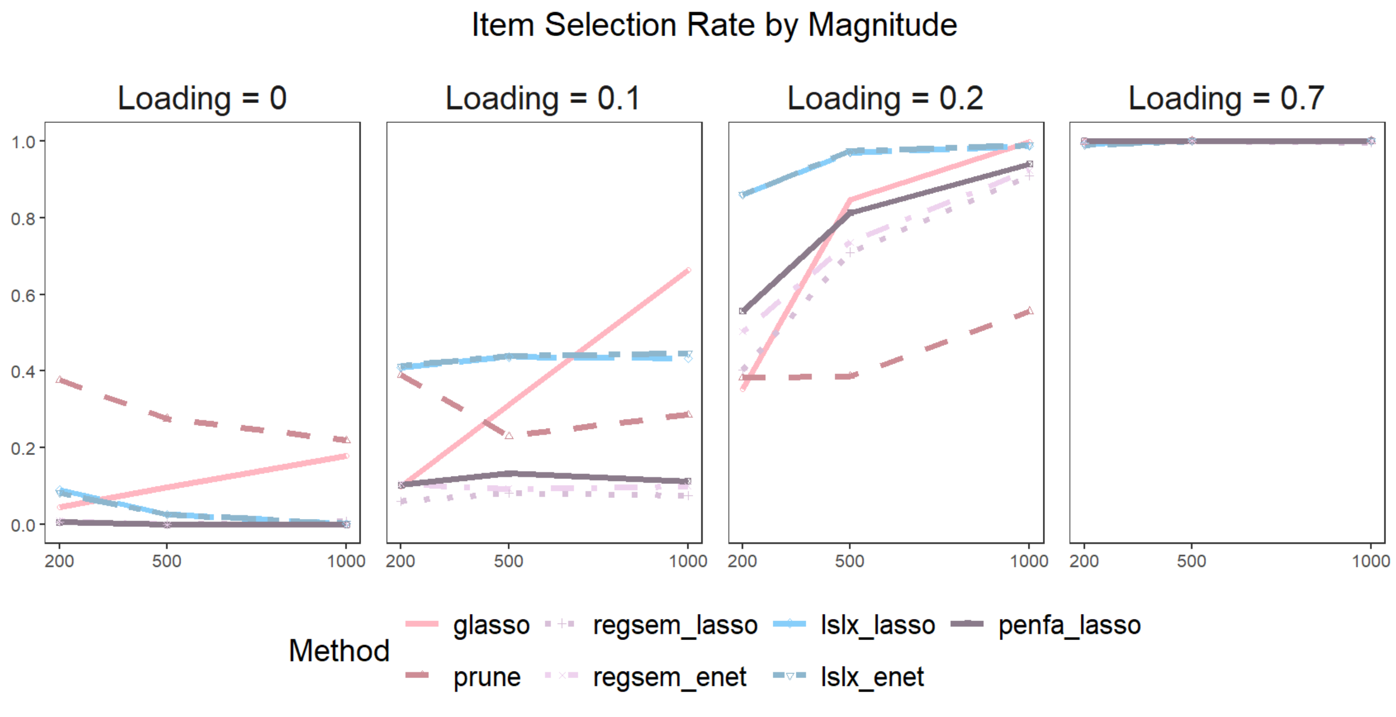
Non-zero items = 21





Results (three factor model with cross-loading)





TP rate = $\frac{TP}{TP+FN}$

FP rate = $\frac{FP}{FP+TN}$

Results (three factor model with cross-loading)

BIC 200

	regsem	penfa	lslx
lasso	12805.89	12763.41	-4.40
enet	12522.44	---	-4.40
	psycho	bootnet	
prune	12972.78	--	
glasso	--	4380.27	

BIC 500

	regsem	penfa	lslx
lasso	31682.76	31648.75	-2.19
enet	31700.66	---	-2.18
	psycho	bootnet	
prune	31920.35	--	
Glasso	--	10199.20	

BIC 1000

	regsem	penfa	lslx
lasso	62130.97	63046.82	-1.25
enet	63154.30	---	-1.24
	psycho	bootnet	
prune	63492.08	--	
glasso	--	19692.90	

For Glasso, the number demonstrate EBIC

Takeaway

Conclusion :

- **Factor analysis** perform reliably in small samples, making them suitable for studies with limited data. However, when sample size increase for multidimensional data, regsem and penfa may prefer to be more conservative than glasso.
- **Network methods** tends to over select more noise variables (items with small loadings), especially for larger samples. Glasso's adaptive regularization allows it to recover more variables as data grows, while fixed-threshold pruning falls behind.
- **Network methods** struggle to detect strong loadings in unidimensional models due to high factor correlation but achieve better performance in multidimensional models.

Future direction :

- Different types of model misspecification or item characteristics

Thank you

Jiaying Chen jc168@uark.edu
Xinya Liang xl014@uark.edu
Jihong Zhang jzhang@uark.edu