EXPLORATORY MEDIATION ANALYSIS WITH MANY POTENTIAL MEDIATORS

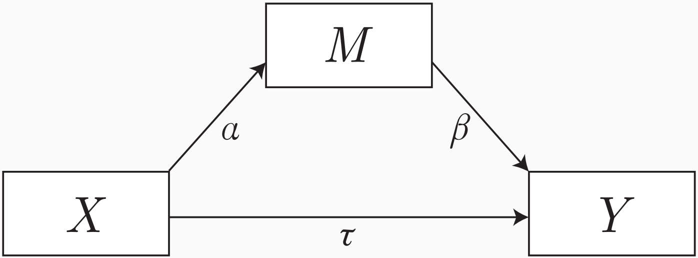
Erik-Jan van Kesteren & Daniel Oberski

April 13, 2018

Utrecht University, Department of Methodology & Statistics

MEDIATION

Q: When is *M* a mediator?



MacKinnon et al. (2002):

1. Causal steps: α & β
2. Difference in coefficients: $\tau - \tau|M$
3. Product of coefficients: $\alpha \times \beta$

VanderWeele (2015, *p.* 46): *"Also take into account $X \cdot M$ interaction!"*

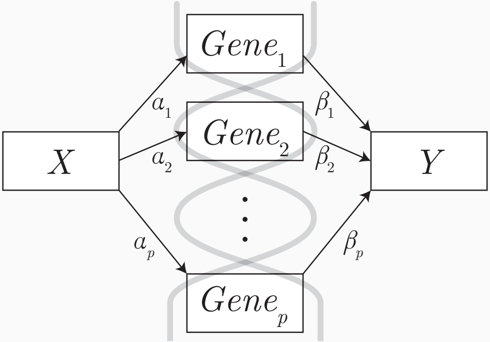
Theory-based decision functions using data from X, M, Y:

$$\mathcal{D}\colon \{\boldsymbol{x}, \boldsymbol{m}, \boldsymbol{y}\} \mapsto \{0, 1\}$$

(0 = not mediator, 1 = mediator)

MANY MEDIATORS

Q: When is $Gene_i$ a mediator?



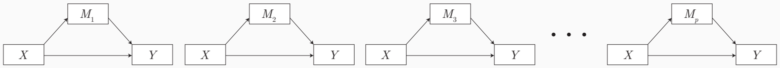
Preacher and Hayes (2008):

1. Fit the full SEM so your parameter estimates take all mediators into account
2. Select mediators using the estimated parameters

$$\mathcal{D}(\boldsymbol{x}, \boldsymbol{m}_i, \boldsymbol{y}) \text{ conditional on } M_{-i}$$

FILTER

p single mediator models



$\texttt{for (i in 1:}p\texttt{)}$ $\mathcal{D}(\boldsymbol{x}, \boldsymbol{m}_i, \boldsymbol{y})$

The “filter” method (Guyon and Elisseeff, 2003)

Good

- Simple
- Quick
- Flexible

Bad

- Assumes uncorrelated mediators: won't work if mediation only visible conditionally

Jacobucci et al. (2016): We can now penalise SEM parameters

$$F_{\text{regsem}} = F_{\text{ML}} + \lambda P(\cdot)$$

Serang et al. (2017): We can use this to select mediators! Put a lasso penalty on α and β

The “XMed” method

Good

- ”Full” SEM
- Does not assume uncorrelated mediators
- Regularisation is hip

Bad

- What are we actually optimising for?
- Find M for which α OR β but we want α AND β.

Our contribution:

$$\mathcal{D}(\boldsymbol{x}, \boldsymbol{m}_i, \boldsymbol{y}) \text{ conditional on } M_{-i}$$

Insight from regularisation literature (Hastie et al., 2015):

conditional parameter == parameter estimated on residual

Idea:

```
sel ← rep(0, p)

while (!convergence) {
  for (i in 1:p) {
    x_res ← x - M[, sel] %*% beta_x_sel
    y_res ← y - M[, sel] %*% beta_y_sel
    sel[i] ← decisionFunction(x_res, M[, i], y_res)
  }
}
```

for each mediator	Coordinate-wise
perform the decision function	Mediation
throw it out if 0	Filter

conditional on the other selected mediators

repeat until convergence

Good

- Uses theoretically relevant $\mathcal{D}$
- Does not assume uncorrelated mediators

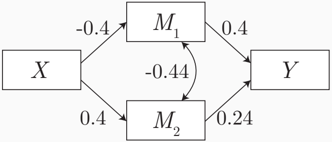
Bad

- Nonconvergence

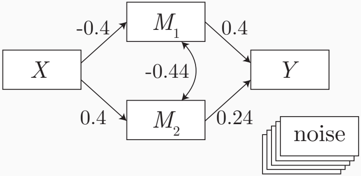
SIMULATION

SIMULATION

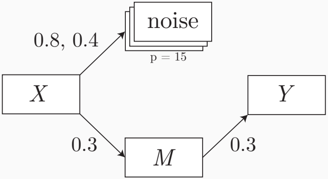
Conditional-only



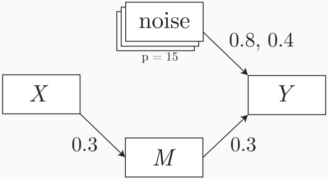
High-dimensional



Noise (α paths)



Noise (β paths)

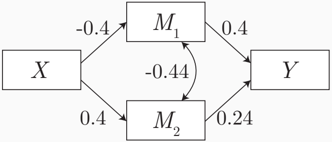


Suppression

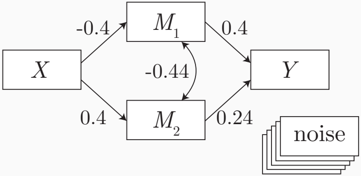
Method	Power M_1	Power M_2
SEM	1	1
Filter	1	0
XMed	1	1
HIMA	1	1
CMF	1	1

SIMULATION

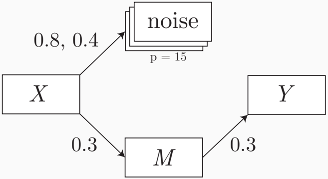
Conditional-only



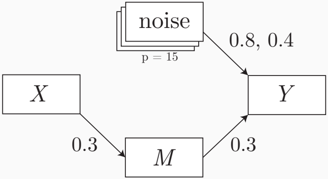
High-dimensional



Noise (α paths)



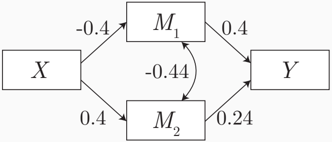
Noise (β paths)



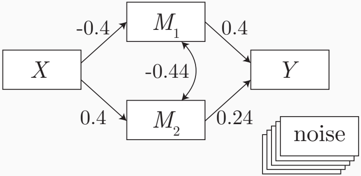
Noise α

	M	.	.	4	.	.	.	.	.	10	.	.	.	14	.	.
SEM	**100**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Filter	**100**	.	.	100	.	.	.	.	.	100	.	.	.	68	.	.
XMed	**77**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HIMA	**100**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CMF	**100**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

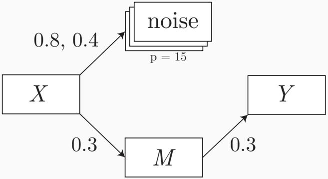
Conditional-only



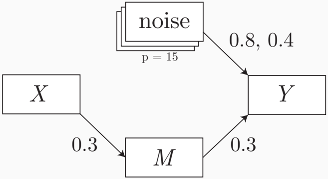
High-dimensional



Noise (α paths)



Noise (β paths)



Noise β

	M	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
SEM	**100**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Filter	**100**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
XMed	**100**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HIMA	**.**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CMF	**100**	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

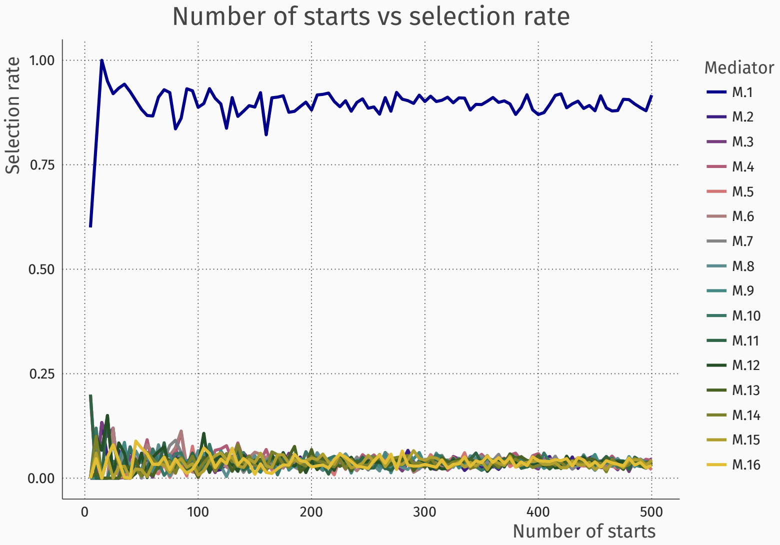
IMPLEMENTATION

HOW TO USE

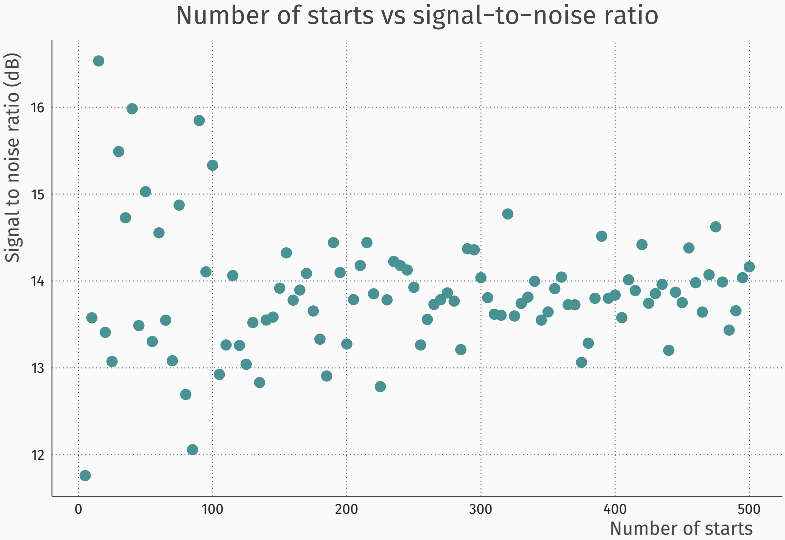
```
> devtools::install_github('vankesteren/cmfilter')
> library(cmfilter)
> res <- cmf(x, M, y, verbose = TRUE)

# CMF Algorithm
#
# -----------
#
# 1 0 0 1 1 1 1 0 1 0 0 0 0 1 1 0 0
# 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
# 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
# 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
#
# Algorithm converged
#
# -----------
```

```
cmf(x, M, y,
    decisionFunction,
    maxIter,
    stableLag,
>>  nStarts,       <<
>>  randomStart,  <<
>>  subSampling,  <<
    parallel,
    nCores,
    progressBar)
```

SUBSAMPLING



SUBSAMPLING



CONCLUSION

- Novel method for selecting among many mediators
- Flexible choice of $\mathcal{D}$
- Conditional on M_{-i}
- Stable in boundary cases
- Traditional power/type-I tradeoff

- Group lasso
- Bayesian regularisation prior on $\alpha \times \beta$
- “Directions of mediation” (Chén et al., 2017)

REFERENCES

Chén, O. Y., Crainiceanu, C., Ogburn, E. L., Caffo, B. S., and Wager, T. O. R. D. (2017). High-dimensional multivariate mediation with application to neuroimaging data. *Biostatistics*, (September):1–16.

Guyon, I. and Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research (JMLR)*, 3(3):1157–1182.

Hastie, T., Tibshirani, R., and Wainwright, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. CRC Press, Boca Raton.

Jacobucci, R., Grimm, K. J., and McArdle, J. J. (2016). Regularized Structural Equation Modeling. *Structural Equation Modeling*, 23(4):555–566.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., and Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological methods*, 7(1):83–104.

Preacher, K. J. and Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40(3):879–891.

Serang, S., Jacobucci, R., Brimhall, K. C., and Grimm, K. J. (2017). Exploratory Mediation Analysis via Regularization. *Structural Equation Modeling*, 24(5):733–744.

VanderWeele, T. J. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press, New York, 1st edition.

QUESTIONS