Assessing Continuous Versus Categorical Diagnosis Using Latent Variable Modeling

Ross Jacobucci
Maxwell Hong
Andreas Brandmaier
Brooke Ammerman
Michael McCloskey

Motivating Example

Psychological Assessment

1040-3590/16/$12.00 http://dx.doi.org/10.1037/pas0000334

Development and Validation of Empirically Derived Frequency Criteria for NSSI Disorder Using Exploratory Data Mining

Brooke A. Ammerman
Temple University

Ross Jacobucci
University of Southern California

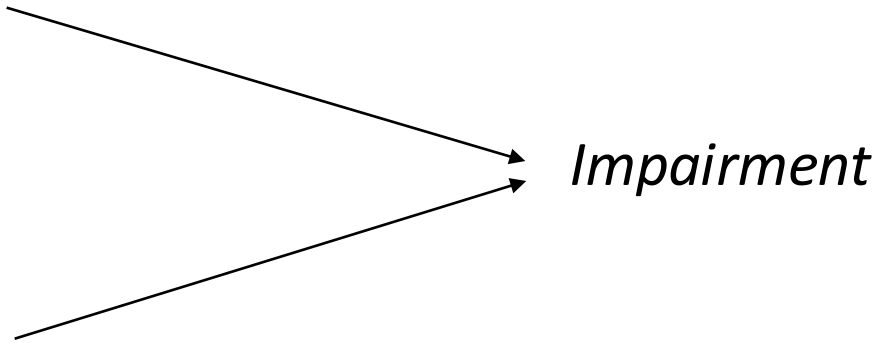
Evan M. Kleiman
Harvard University

Jennifer J. Muehlenkamp
University of Wisconsin, Eau Claire

Michael S. McCloskey
Temple University

Goal of Analysis

- To find clinically meaningful sub groups of non-suicidal self-injury (NSSI)
  - Cutoffs for DSM-V Criteria
  - Current cutoff = 5 NSSI acts in past year
    - Workgroup for disorder stated this is an "arbitrary cutoff"
- Have measure of how many times participants self-injured in last year
  - "Have you ever, intentionally or on purpose, hurt yourself in the following ways, without the intention of killing yourself?"
- Outcome
  - One-factor Factor Analysis consisting of:
    - Suicidality
    - Emotion Dysregulation
    - Emotion reactivity
    - Borderline personality
    - Disordered eating
    - Anxiety
    - Depression



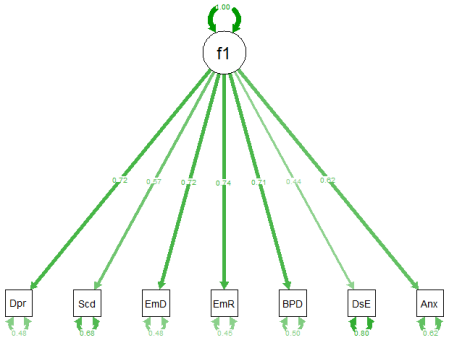
Proposal

Necessary to Specify Two Parts

- 1. Outcome of interest
  - What will the resultant subgroups present differences on?
- 2. Variables that define subgroups
  - What variables do we assess for splitting cases?

Outcome

- In the case of our example, an *impairment* latent variable



- Could propose a more complex SEM
  - Other project: Four latent variables – BPD, Trauma, Suicide Behavior, Depression
- This is the confirmatory piece of the model
  - Requires the strongest theoretical rationale

Outcome Continued

- The results only hold given the sample we used
- Preferable to perform an Integrative Data Analysis (e.g. Bauer & Hussong, 2009)
  - Do not necessarily need all outcome variables in each dataset
    - Do need all predictors
- Challenge will be dealing with indicators of the latent variable that are missing in some samples
  - Or worded differently

Predictors

- Variables used to define the subgroups
  - Test for cutoffs, or just as a linear predictor (continuous relationship)
  - In relation to the outcome
- In our case, we have number of recent NSSI acts
- With depression, we could test:
  - Symptoms – Is 5 the right number?
    - Do some symptoms result in more severe scores on the outcome?
  - Include depression scale items (e.g. CESD)
    - Which items, and what response options, result in worse outcome scores?

Predictors

- This is the exploratory piece of the model
  - However, could specify a theoretical model
    - A multiple group model for NSSI using the current cutoffs
    - Only using this model would be extremely limiting
      - The alternative hypothesis would be no groups, not better defined groups
- Worth “controlling” for certain variables
  - Race, ethnicity, sex, etc…
  - Easy to understand in linear framework
  - Create conditional splits in a tree model
    - E.g. cutoff should be at 5.5 for males, 7.5 for females

Categorical vs. Continuous

- There are multiple options for functional form of the relationship between predictor(s) and outcome
  - Each corresponds to a different theoretical conclusion
- In our paper, we only considered groups vs. no groups
  - But what if a linear relationship is more appropriate?
    - E.g. each increase in NSSI results in the same change in impairment?
  - This would provide support for a continuous (dimensional) disorder.
    - Meaning the creation of cutoffs would result in a loss of information

Framework Proposal

- Specify a multivariate outcome of interest
  - Any form of SEM
  - Theory based
- Specify the predictors, or what variables directly account for the criteria
  - These are used to create groups and test the relationship form
  - Does not need to be theory based
- Perform a model comparison approach
  - Identify whether cutoffs are appropriate or not
  - Compare a tree model, linear MIMIC, and linear spline MIMIC
  - Also test the inclusion of covariates
    - How this is incorporated depends on theory

SEM Trees Overview

When We Can’t Reduce the Outcome to a Single Value

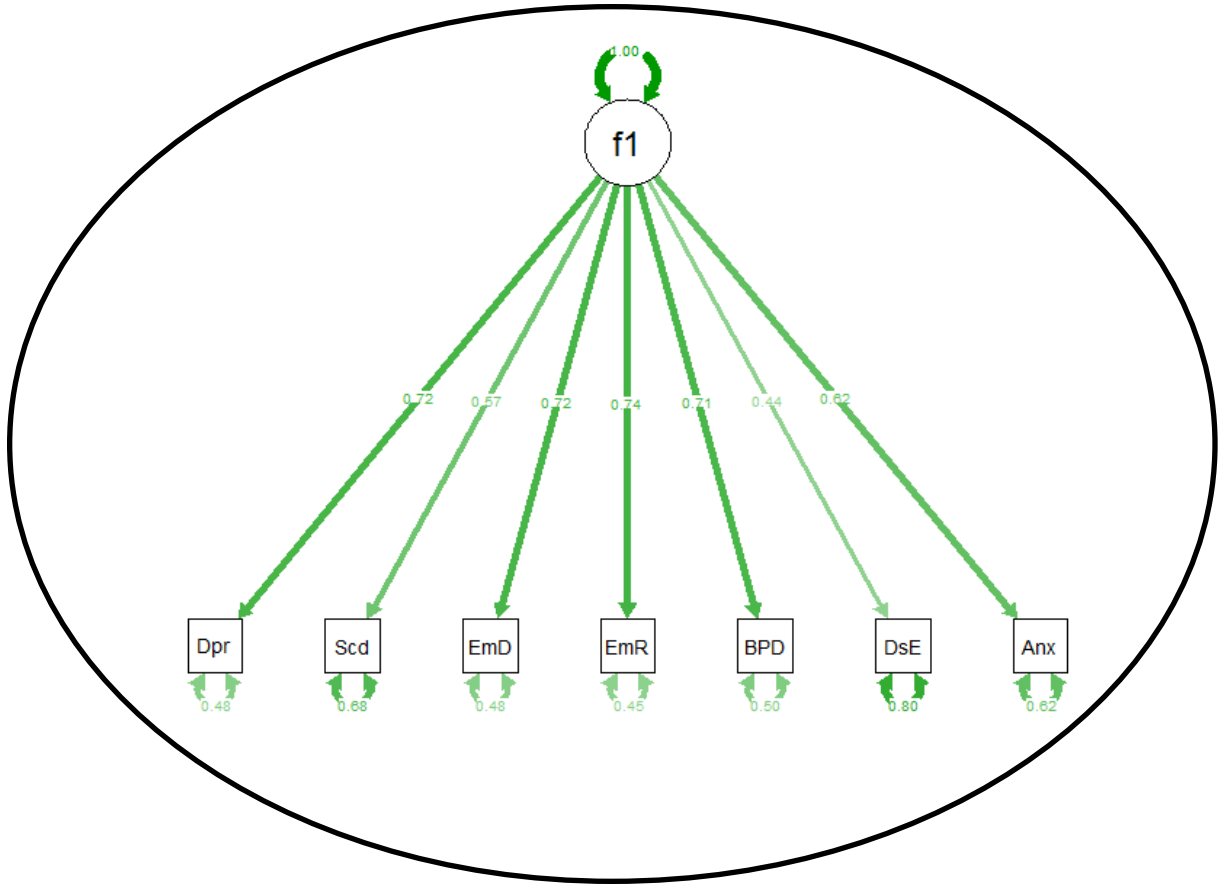
- Need SEM Trees or other multivariate model (Multivariate Boosting or others)
- Instead of reducing our outcome to a single variable we can use:
  - Means of multiple variables
  - A confirmatory factor model
  - Latent growth model
  - Autoregressive model
  - Etc…

SEM Trees

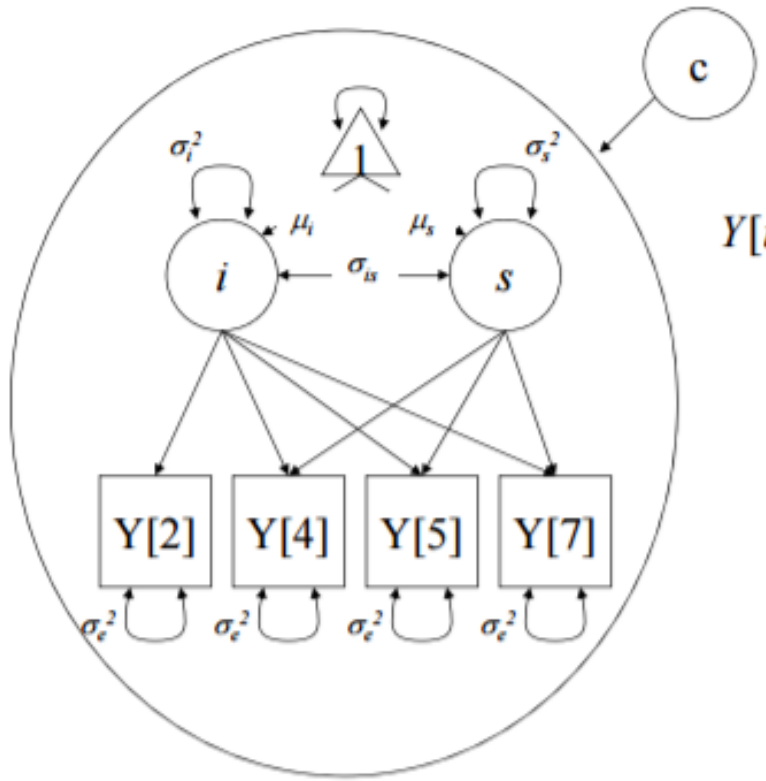
Predictor

Outcome

NSSI Acts



Growth Mixture Model



$$Y[t]_n = \sum_{k=1}^{K} \pi_{nk}(i_n + s_n \cdot A[t] + e[t]_n)$$

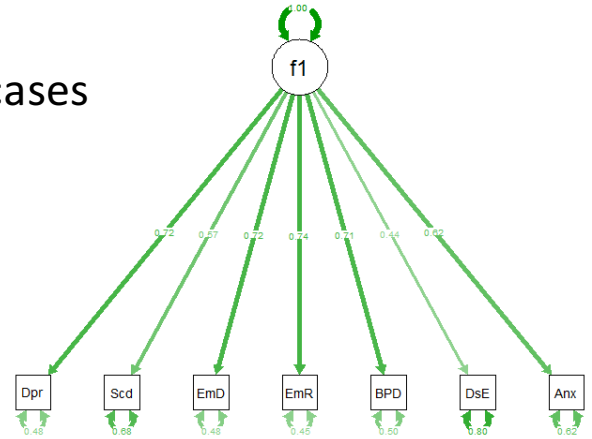
SEM Tree Cont'd

- Different than just including a covariate in the model
- In SEM Trees, the covariates predict the *model fit*
  - Not just the latent variable
- In predicting the fit of the model, you are indirectly predicting differences in each of the model parameters
  - i.e. factor variance, mean, loadings ….
  - In other words: If you change the model parameters then you change the model fit
- This is an ***exploratory*** search for multiple group models
  - Jacobucci, Grimm, & McArdle (2017)

SEM Trees Algorithm Cont'd

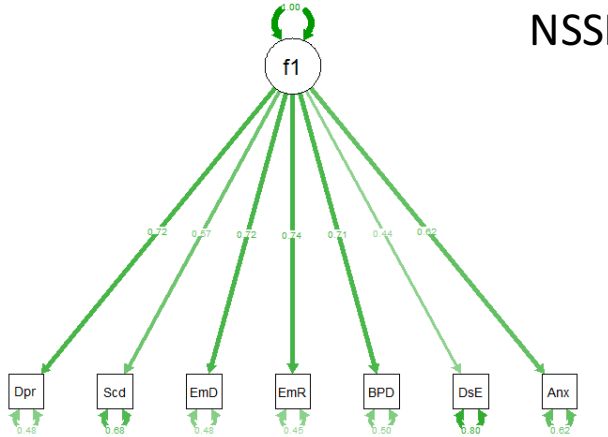
- At each tested split, the model becomes

All cases



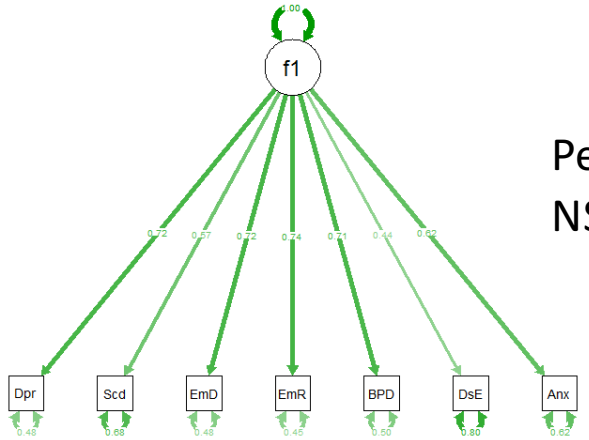
VERSUS

People w/
NSSI = 0



+

People w/
NSSI > 0

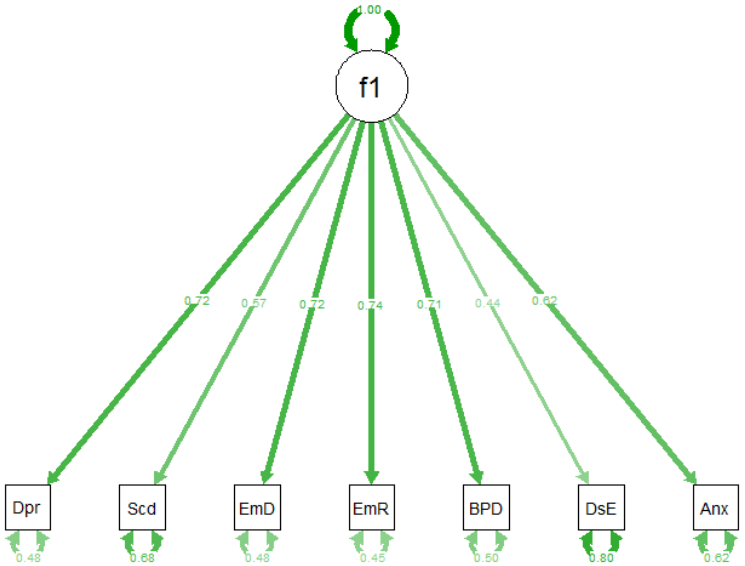


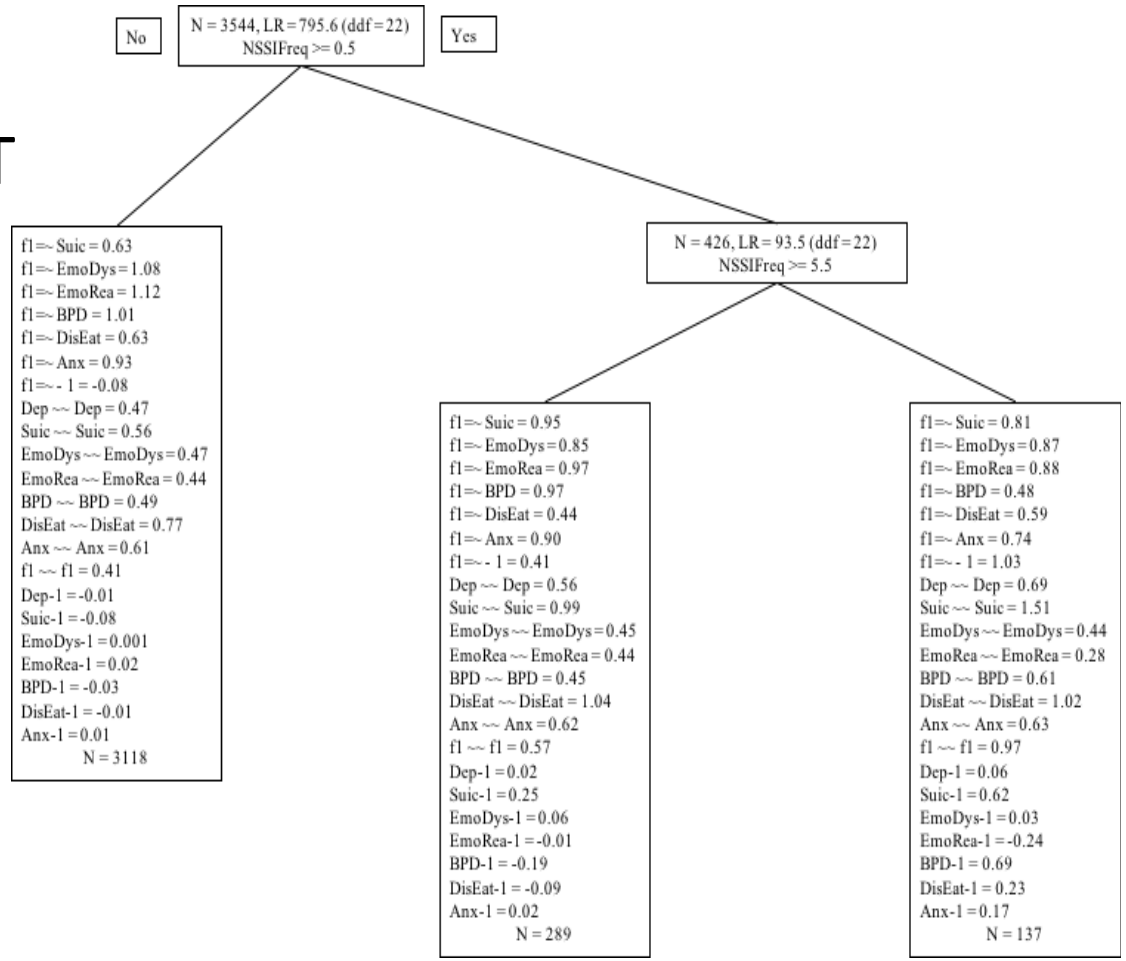
SEM Trees Algorithm Cont'd

- If the covariate is an integer (& ordinal), tests every possible group in sequence
  - 0 vs. 1-1000, 0-1 vs 2-1000, 0-2 vs 3-1000 etc…
- Same thing if numeric (continuous; e.g. 0.324)
  - Will test every value in sequence, so may be better to round first
- If categorical (factor), one vs. the rest scheme
  - If 5 categories, 15 possible splits
    - Computationally intensive
- Once a best split is determined:
  - Move down one level to start over searching for an additional split
  - Continues until:
    - No longer improves model fit
    - Reaches other stopping criterion
      - Too small N in a node
      - A priori set maximum depth (# of groups)

SEM Trees & NSSI

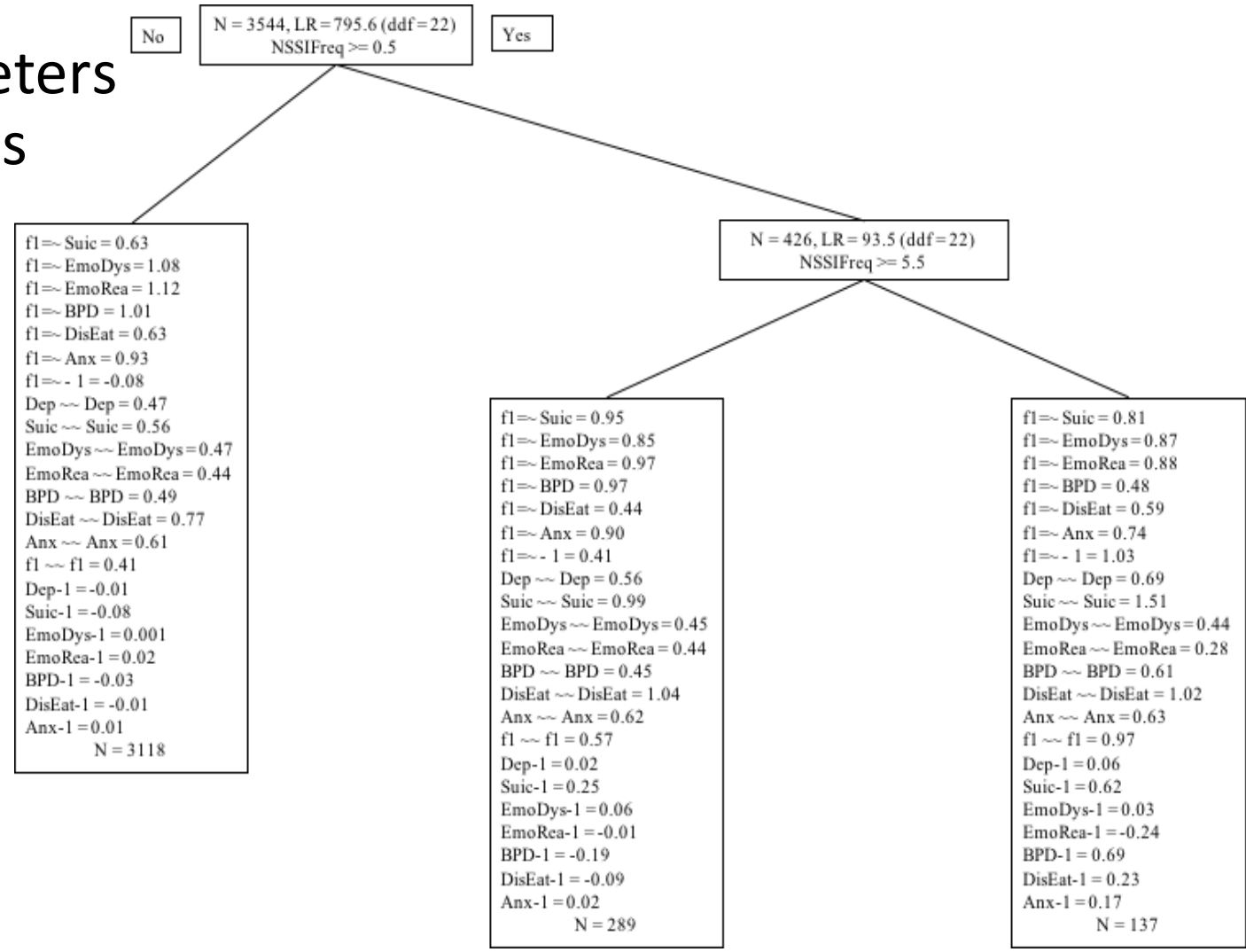
- Use a one factor CFA as the outcome
- Came up with the same splits as with DT





SEM Trees & NSSI Cont'd

- Can investigate individual parameters to get a better clue to how groups differ
  - Most variable means increased
    - BPD & Suicide most
  - Factor variance increased



SEM Trees & NSSI Conclusion

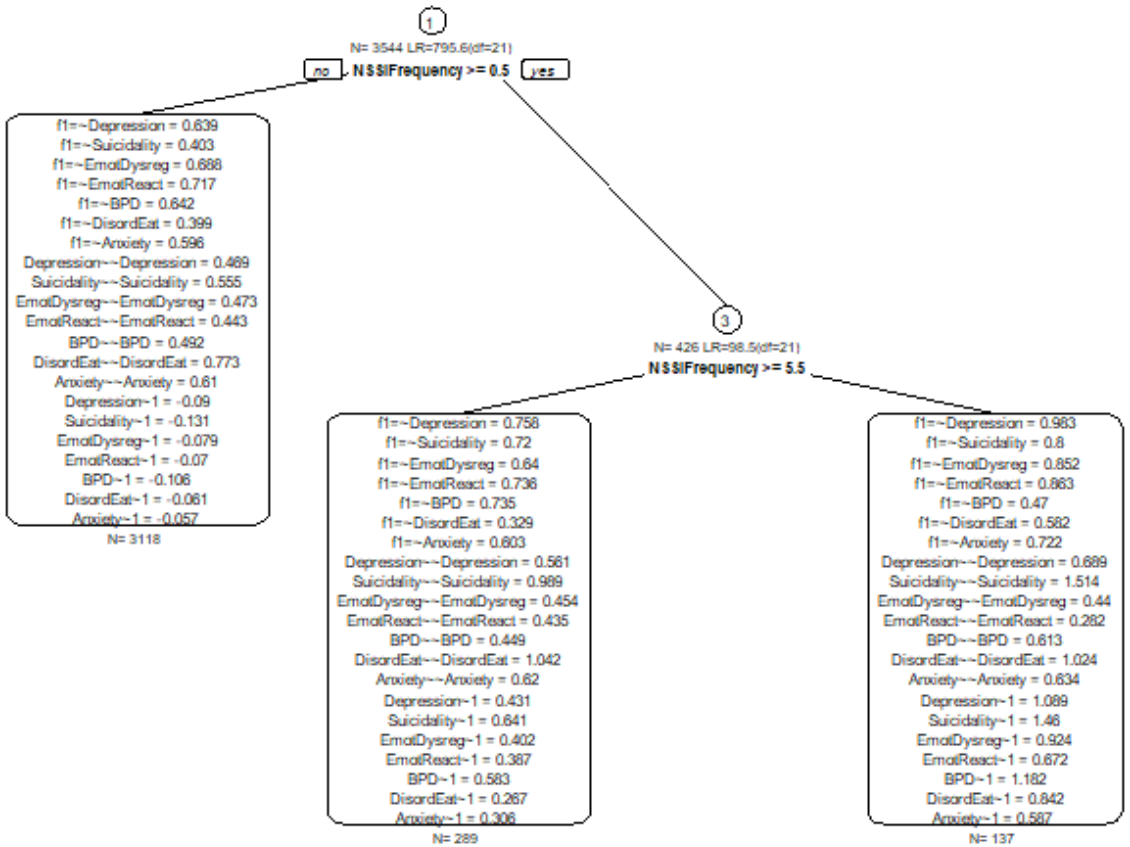
- Agreement in cutoffs across methods
  - Important, as SEM Trees hasn't been studied that much
- SEM Trees allows for a more comprehensive evaluation
  - While being able to make specific comparisons across groups
- Was able to incorporate more information into the model than if we used mixture models
  - This is not always the case in comparing mixtures and SEM Trees
  - We had a ***very*** informative covariate

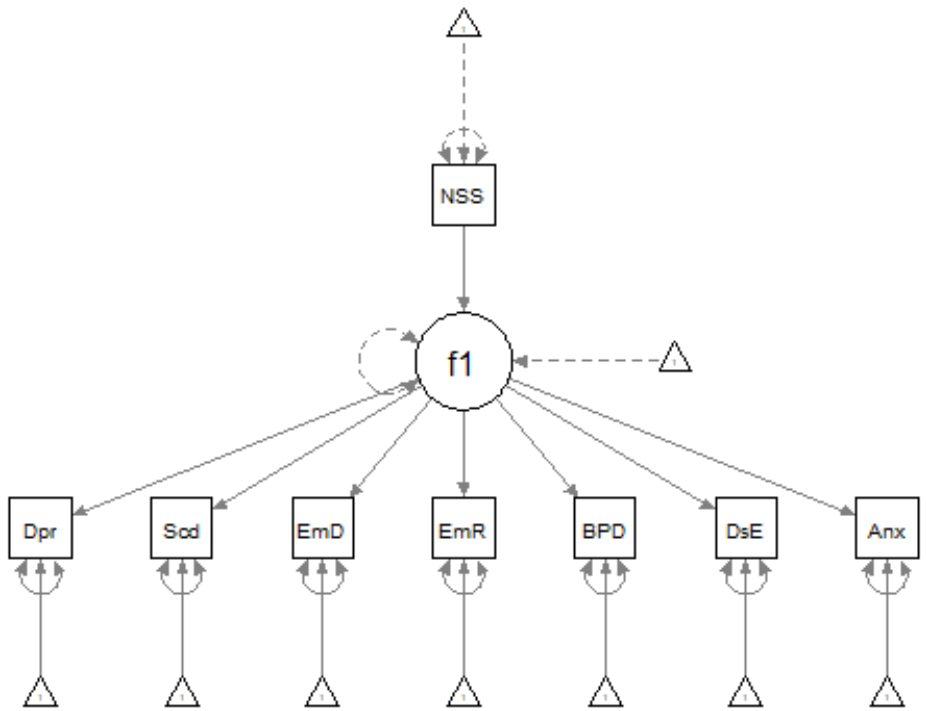
Categorical versus Continuous

What's the Alternative Hypothesis?

- We know there is a relationship between NSSI and general impairment
- Only using SEM Trees could mask a linear relationship
  - **i.e. there aren't subgroups**
  - **Each additional NSSI act results in the same increase (decrease) in impairment**
- **The null hypothesis in SEM Trees is there is no relationship between predictor and log-likelihood**
- Need a way to compare a SEM Trees and linear MIMIC model to disentangle

Model Comparison

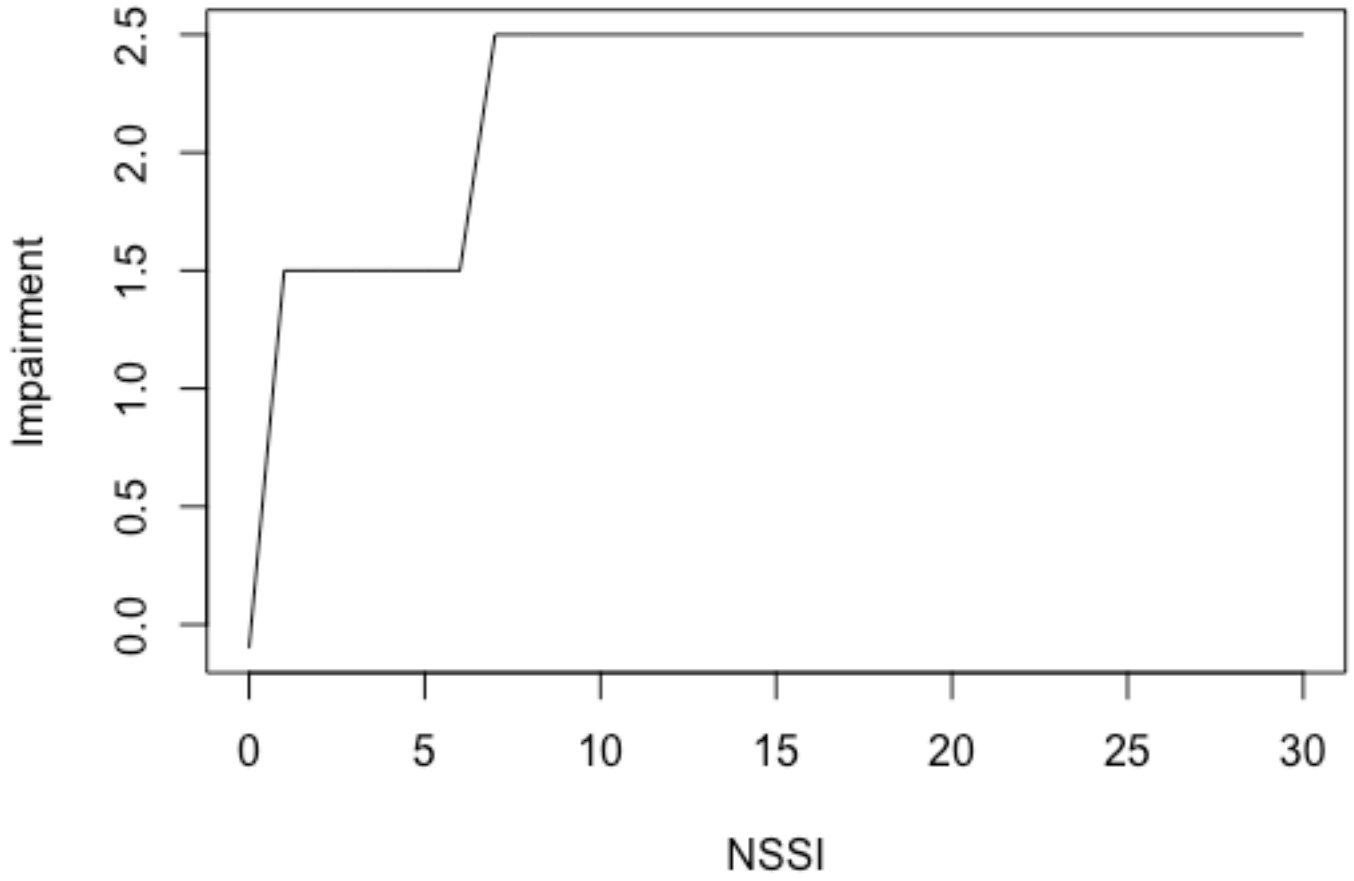




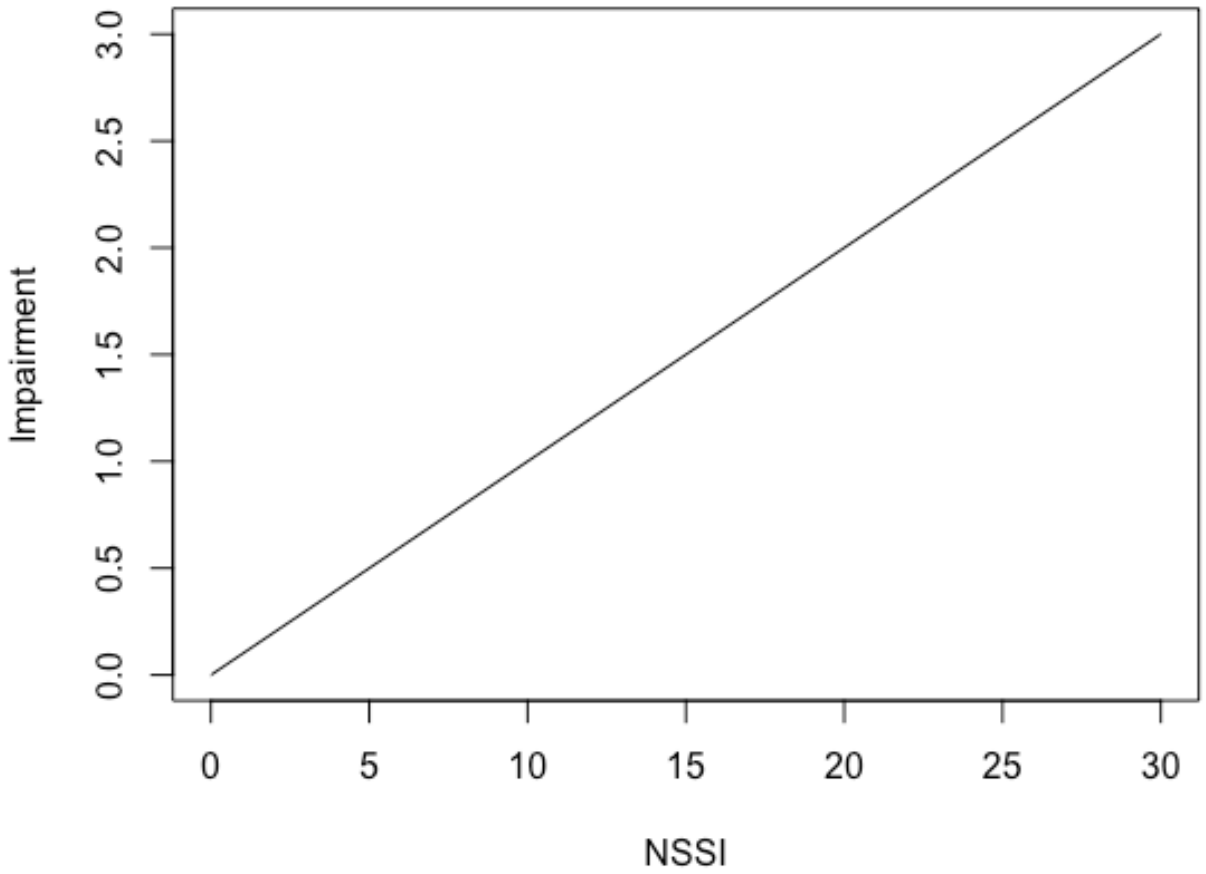
Best Model Fit Implies Functional Form

- With a single outcome, I could perform both DT and linear regression
  - Best fitting model (lowest RMSE using CV) should indicate a more appropriate functional form
    - Similar idea to what we term *Deductive Data Mining* (Hong, Jacobucci, & Lubke, in prep)
- With the factor score as an outcome and using bootstrapping:
  - Linear regression: $R^2 = 0.01$
  - Linear regression w/ sqrt(NSSI): $R^2 = 0.05$
  - Decision Tree: $R^2 = 0.09$
  - MARS: $R^2 = 0.07$
  - Boosting: $R^2 = 0.10$

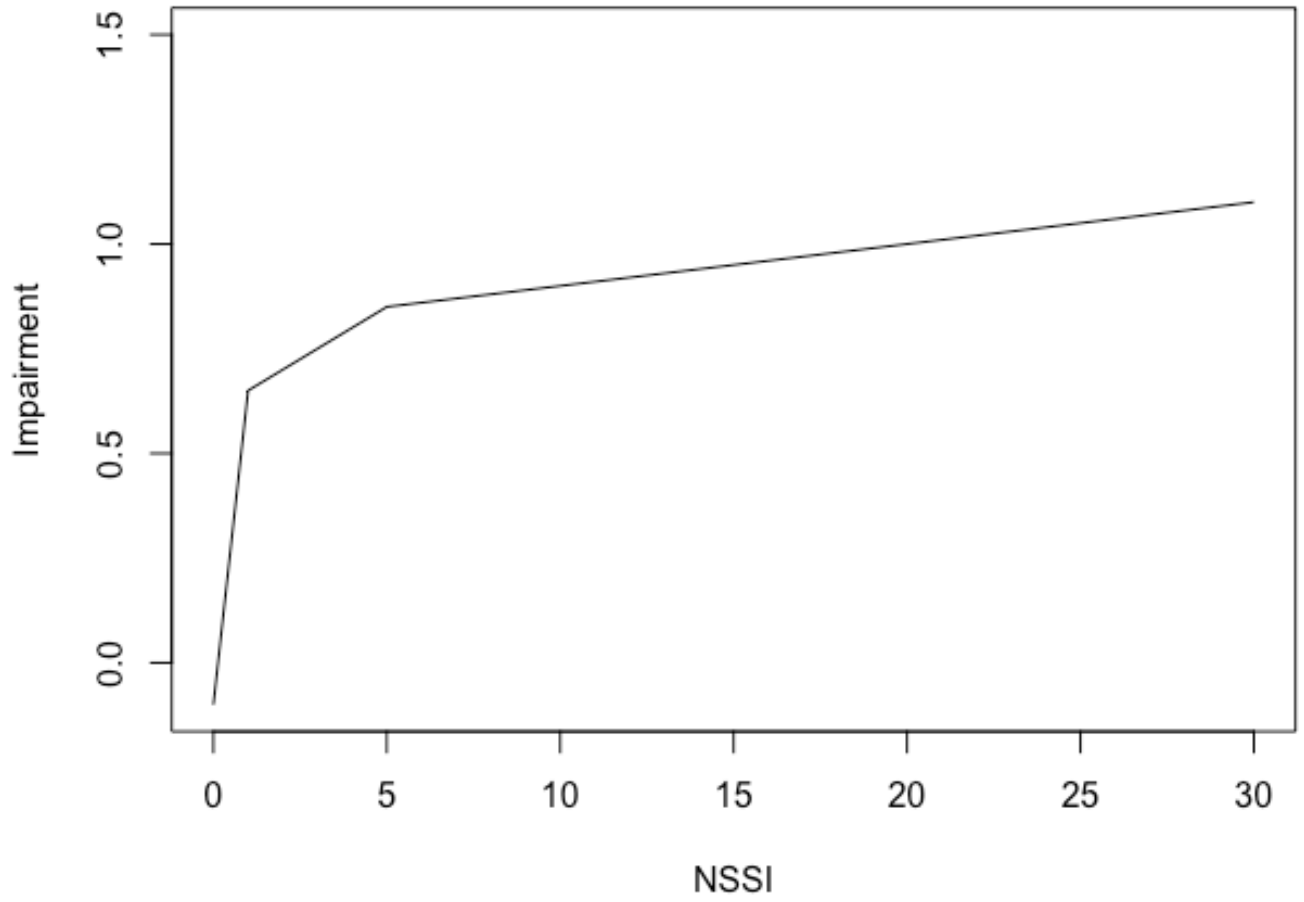
SEM Trees Fundamental Assumption



Linear SEM Fundamental Assumption

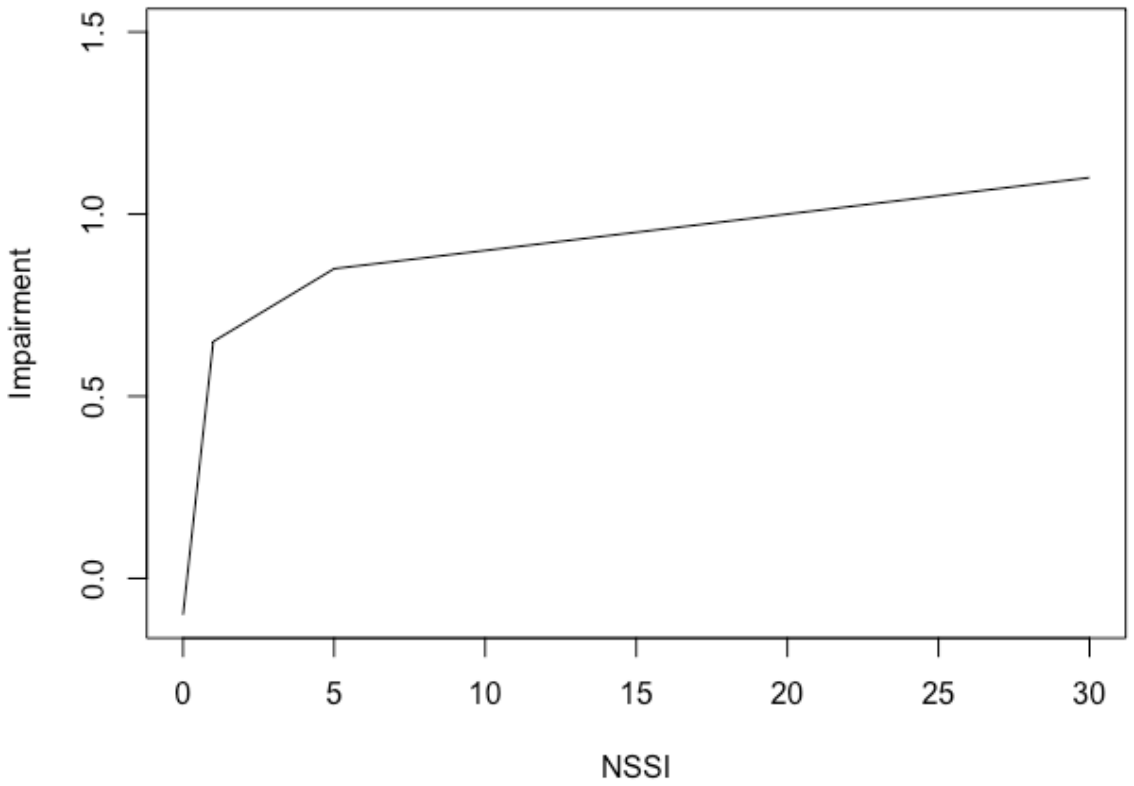


What I Want To Test



Translating to Models

- This last diagram would correspond to a spline/MARS model
  - Knot points indicate cutoffs for groups
- Ways to do this with factor loadings
  - E.g. Longitudinal models
    - (e.g. Grimm or Harring)
  - Doesn't directly translate
    - Split time is a parameter



MIMIC Spline SEM

MIMIC Spline Model

- Expectation for the mean of the latent variable (impairment) is conditional upon NSSI
  - Multiple Indicator Multiple Causes (MIMIC) Model
  - Slope of linear relationship also conditional upon NSSI

$$impair_i = \beta_1 NSSI_i,\ 0 \leq NSSI_i < knot_1$$

$$impair_i = \beta_2 NSSI_i,\ knot_1 \leq NSSI_i < knot_2$$

$$impair_i = \beta_3 NSSI_i,\ knot_2 < NSSI_i \leq 50$$

Frequentist MIMIC Spline

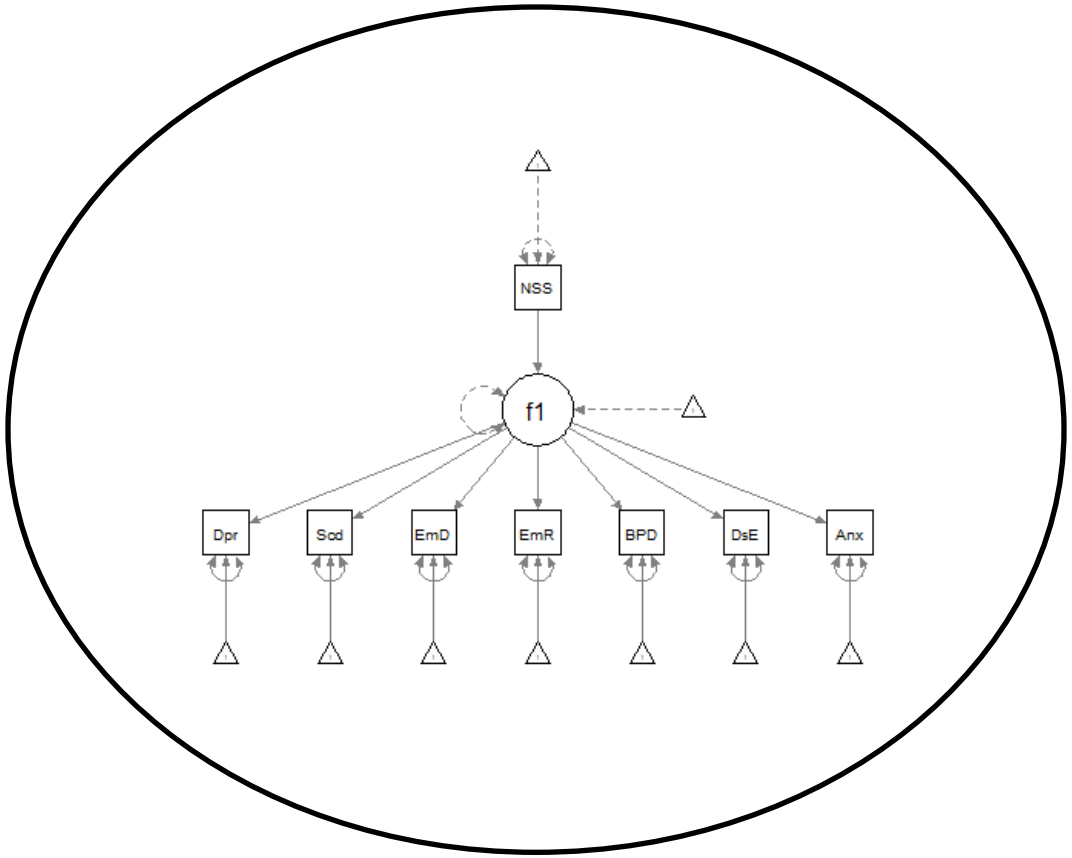
- Don’t think fully possible in any SEM software
  - Maybe OpenMx and/or Mplus?
- In Mplus, we can do a manual search for optimal knot points
  - Select knot point based on likelihood curve
    - Known as profile likelihood approach (Hall, Lipton, Sliwinski, & Stewart, 2000; McArdle & Wang, 2008)
- Or, we can trick SEM Trees to do this for us
  - Outcome model is a linear MIMIC model
  - Create a copy of the predictor variable(s)

SEM Trees Spline

Predictor

NSSI Acts

Outcome



In this, we can constrain all parameters but the regression – **True spline model**

Or, allow all model parameters to vary – **Integrated model** that contains all other models as sub-models

Proposal Steps

- Compare cut points and linear at each possible grouping
  - E.g. first test one split model (2 groups) versus linear
  - Then if split, retest at each node
  - If split fits best, test the stability
- In a disorder such as NSSI, 0 versus 1+ acts are very different groups
  - Don’t really need stats to tell you this
- Allows you to find a 0 versus 1+ split, but that there is a linear increase among those with 1+
- Really is a way have an alternative solution for each possible grouping
  - If a linear relationship, SEM Trees would split because this is still better than no split

Issues in this Model Comparison

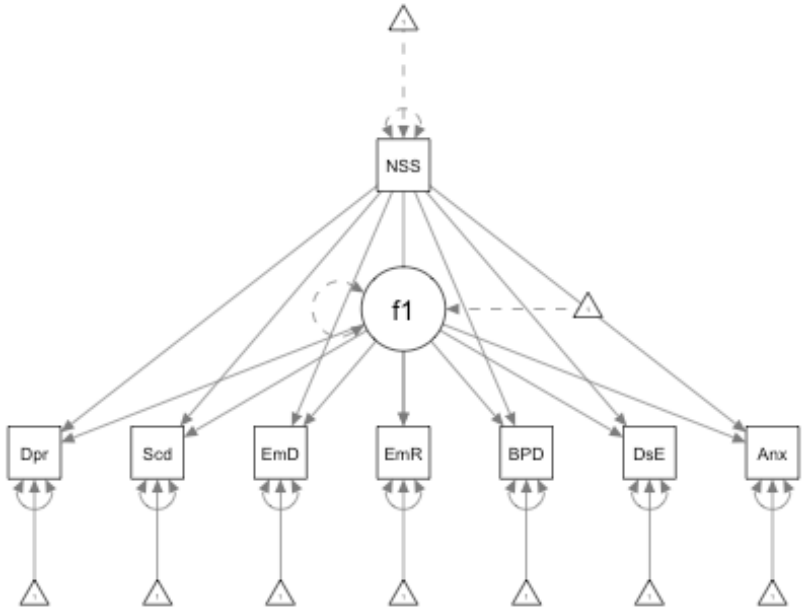
Issue #1: Number of parameters

- Tree Construction Should be done with no invariance constraints
  - Unless some a priori idea, it is worthwhile assessing what parameters could be driving the group differences
    - Factor loadings, residuals, latent variance, etc.
    - Allows a fairer comparison between MIMIC and Multiple Group on BIC
- Brings up a separate point: Is invariance necessary?
  - What if I find groups that differ with respect to differing constructs?
    - E.g. each group has very different factor loadings

Issue #2: MIMIC Assumptions

- If you work out expectations, the effect of a predictor on each manifest variable is a product of the regression coefficient and factor loading
  - This could mask specific effects
- Worth testing a model such as:
  - Or saturated covariance

    as outcome



Issue #3: Cut Points

- The use of cut points in decision trees is notoriously unstable (e.g. Breiman, 2001)
- Instability will be more pronounced if a linear model is more appropriate
- Not a good way to overcome this other than actually assessing stability
  - E.g. Jacobucci, under review
- For this, we can use bootstrap sampling to get a descriptive understanding of what cut points are most likely and the variability of these
- Not necessary to get the exact same split, but should be symmetric with that split at the center
- Can also perform with knot points
  - Instability = less appropriate for data

Results

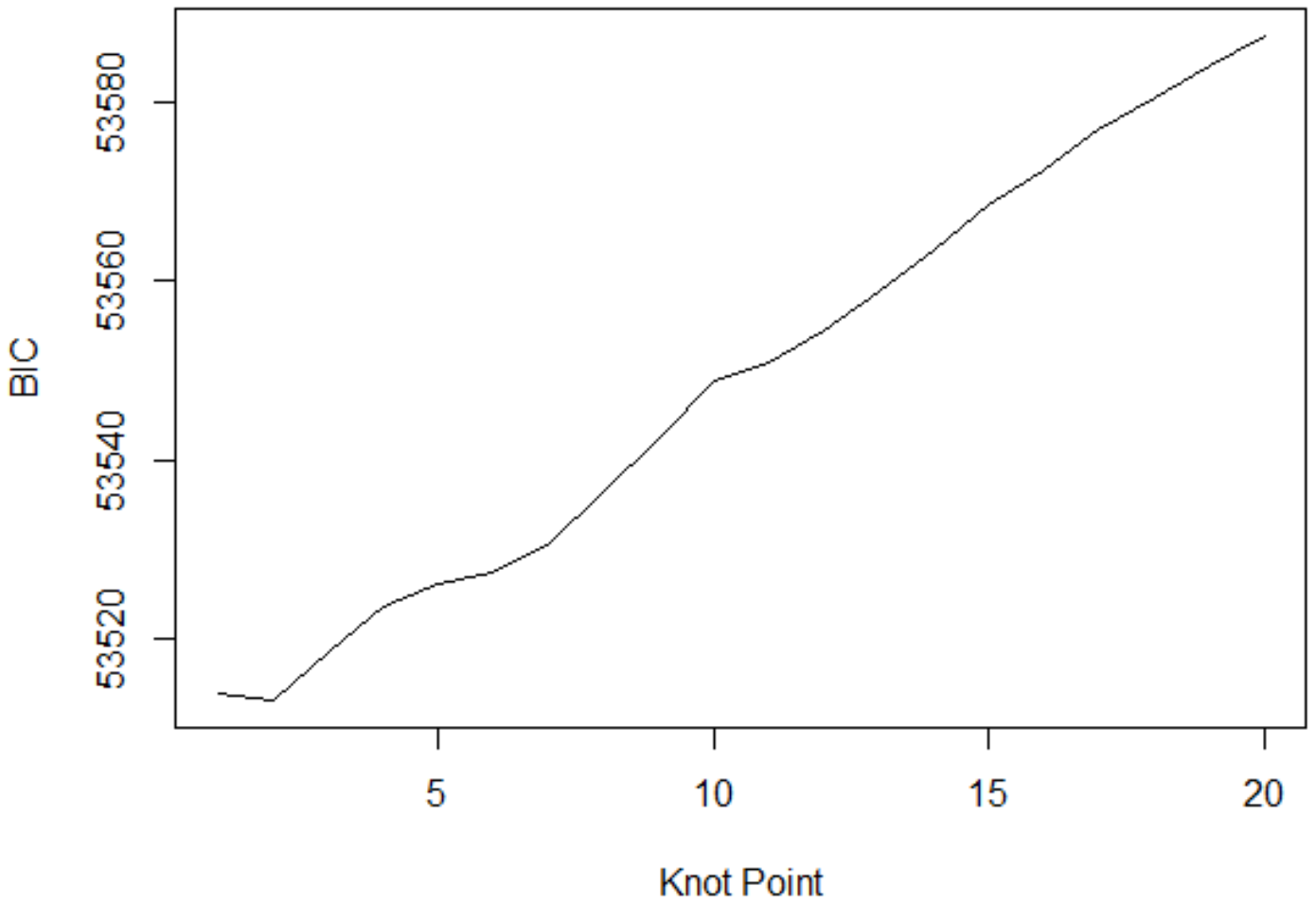
Model Comparison #1: Whole Sample Analysis

- SEM Trees found a split between values 0 and 1

Model	Tree Configural	Tree Metric	Tree Scalar	Tree Strict	Linear MIMIC	Null Linear	Linear Manifest	Linear MIMIC Spline
BIC	**53260**	**53259**	53352	53544	53678	53884	53577	53514

Best fitting model indicates group differences on the impairment latent variable

Spline – Profile Likelihood



Stability of the split at 0.5

Cut Point	0.5	1.5
Count	**45**	5

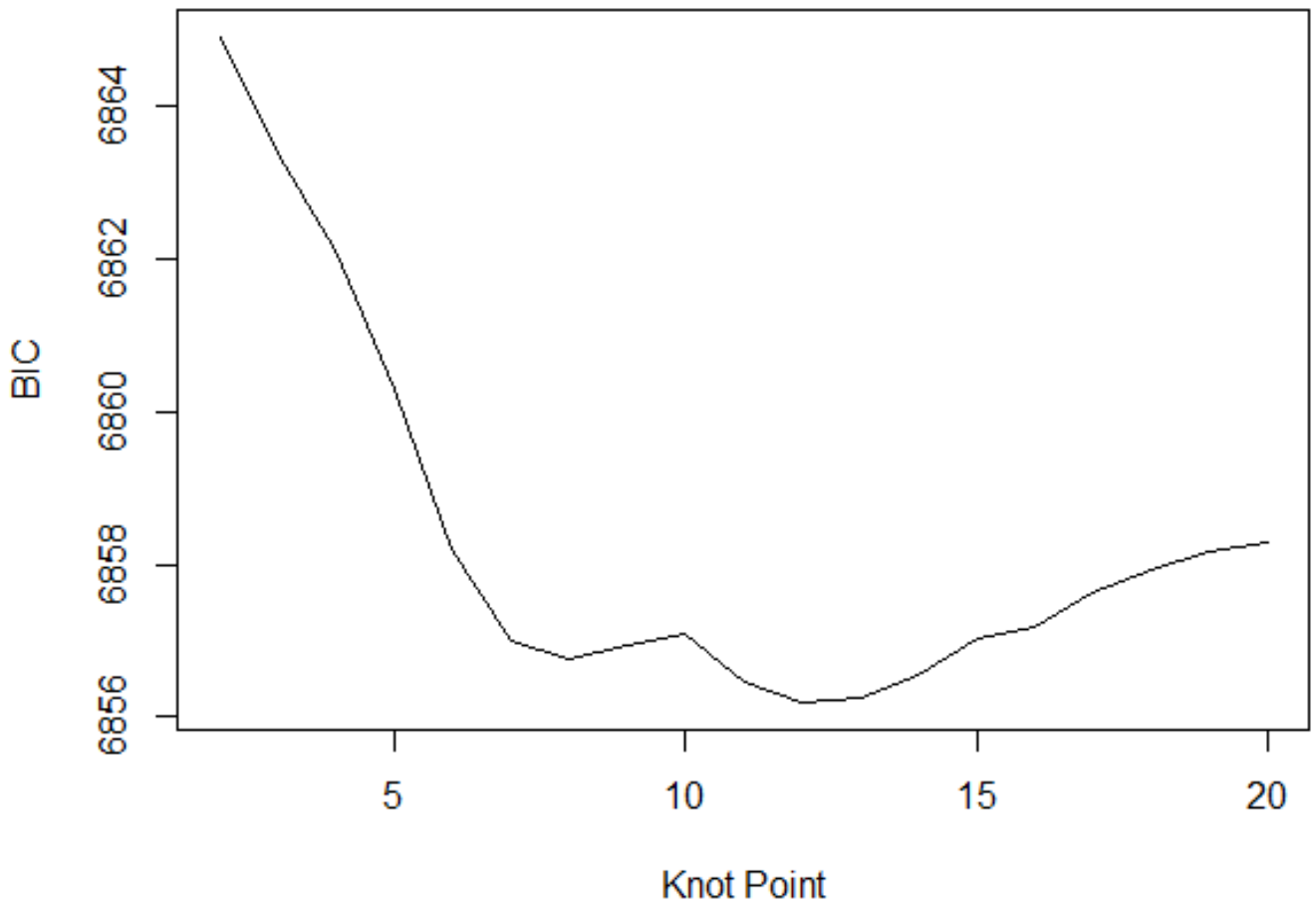
Model Comparison #2: Observations > 0 Acts

- SEM Trees found a split between values 5 and 6

Model	Tree Configural	Tree Metric	Tree Scalar	Tree Strict	Linear MIMIC	Null Linear	Linear Manifest	Linear MIMIC Spline
BIC	6911	6888	6879	**6850**	6862	6882	6870	6856

Best fitting model indicates group differences on the impairment latent variable

Spline



Stability of the split at 5.5

Cut Point	3.5	4.5	5.5	6.5	7.5	8.5	9.5	16.5	19.5	55
Count	3	4	**28**	6	3	1	1	1	2	1

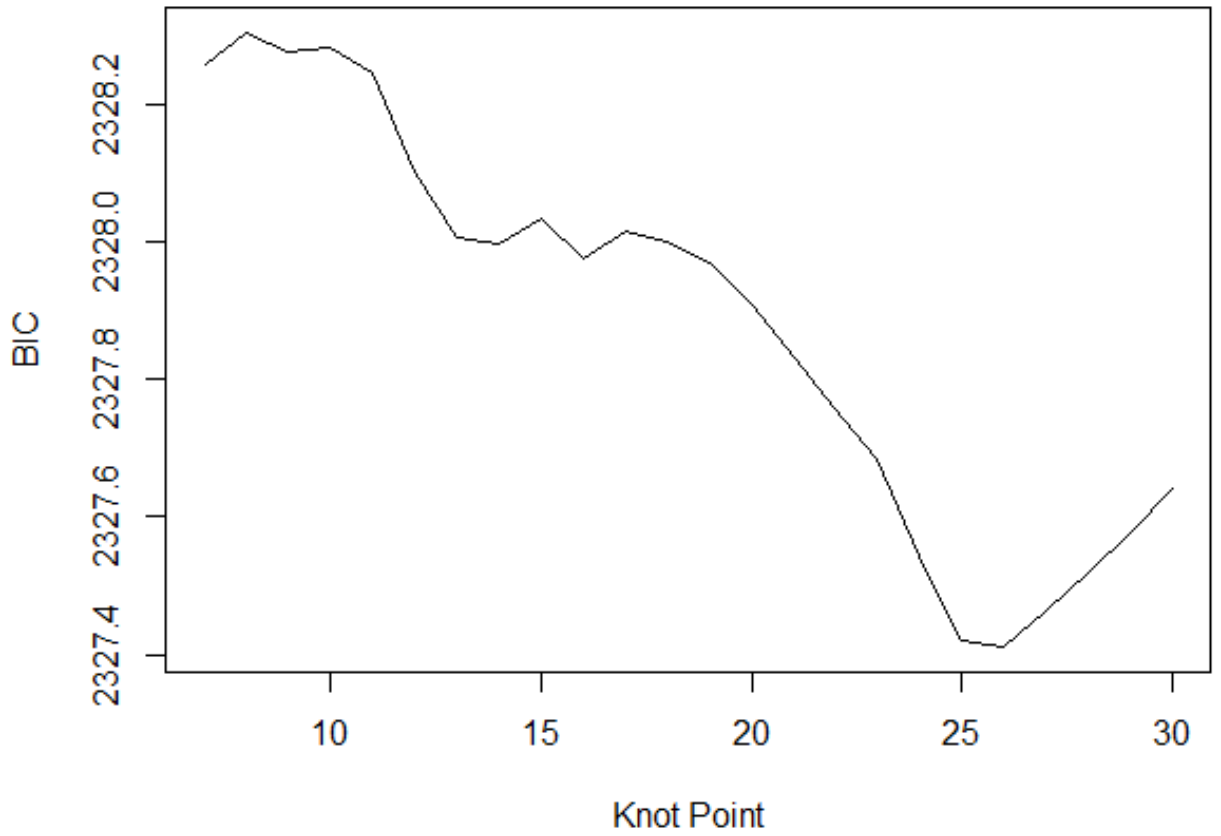
Model Comparison #3: Observations > 5 Acts

- SEM Trees found a split between values 16 & 17

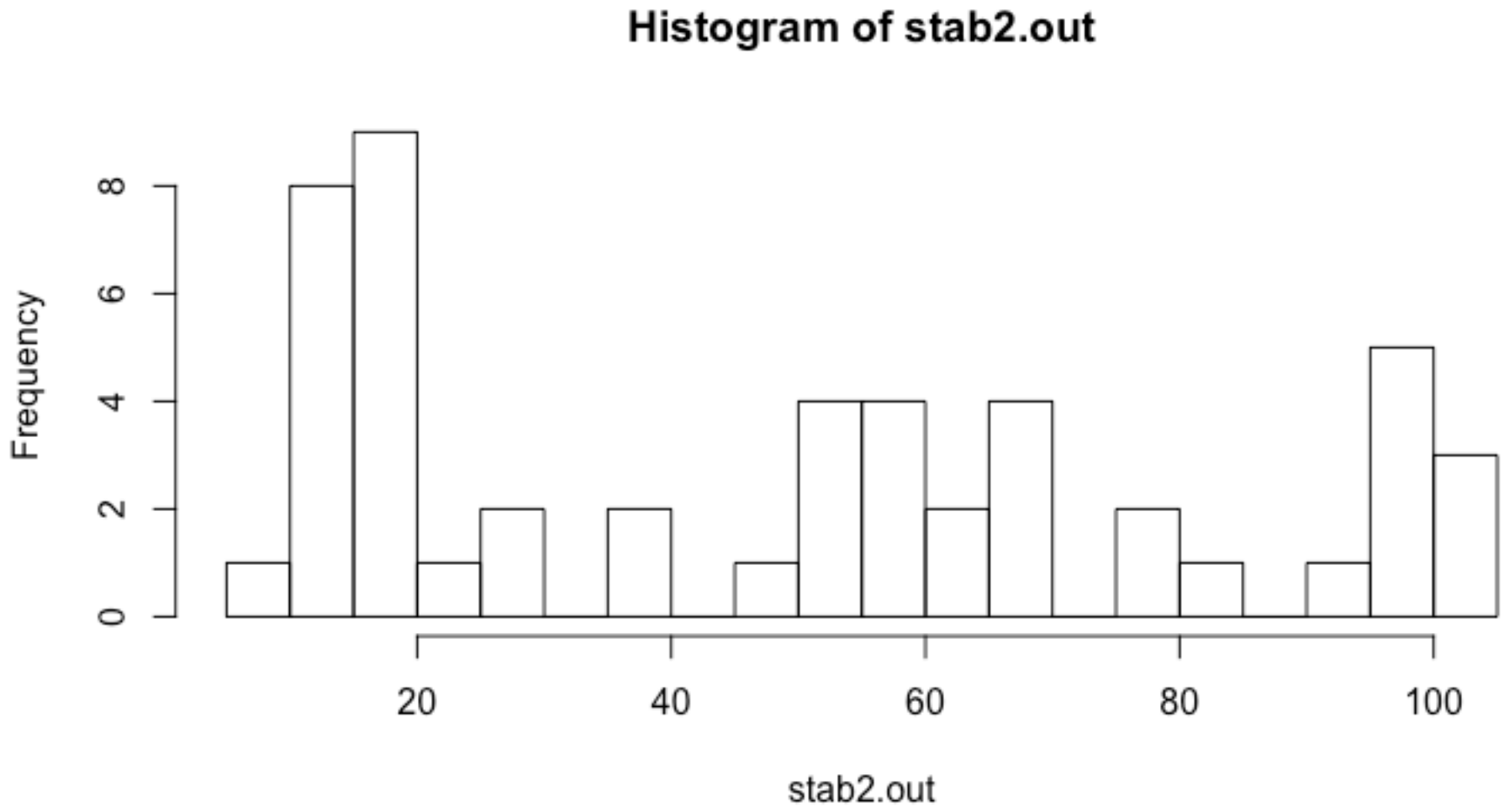
Model	Tree Configural	Tree Metric	Tree Scalar	Tree Strict	Linear MIMIC	Null Linear	Linear Manifest	Linear Spline
BIC	2382	2353	2338	2324	2323	**2319**	2338	2327

Best fitting model indicates no change in impairment as # of acts increases (homogeneity)

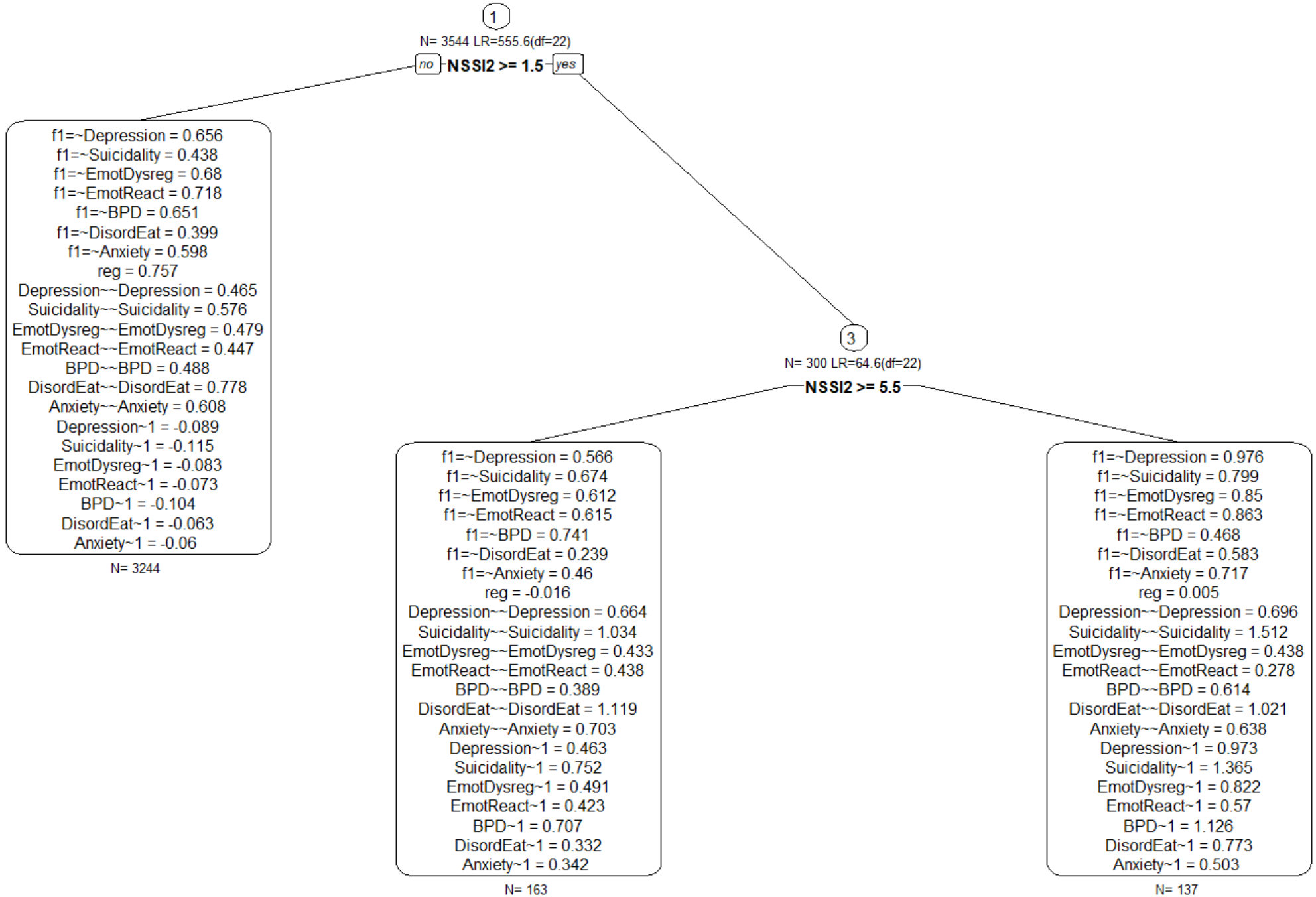
Linear MIMIC Spline



Stability of the split at 16.5



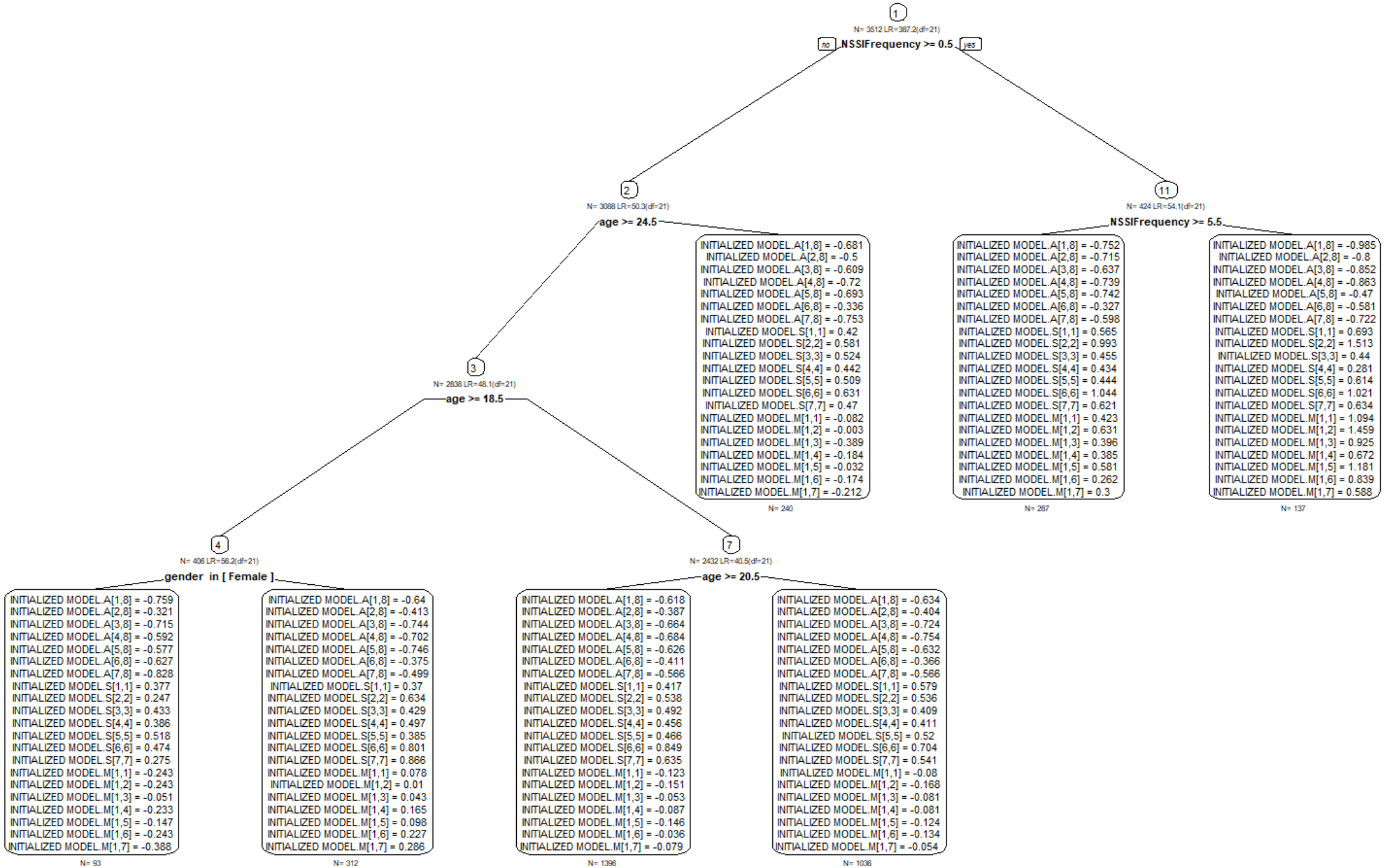
Integrated Model


1
N= 3544 LR=555.6(df=22)
no
NSSI2 >= 1.5
yes
f1=~Depression = 0.656
f1=~Suicidality = 0.438
f1=~EmotDysreg = 0.68
f1=~EmotReact = 0.718
f1=~BPD = 0.651
f1=~DisordEat = 0.399
f1=~Anxiety = 0.598
reg = 0.757
Depression~~Depression = 0.465
Suicidality~~Suicidality = 0.576
EmotDysreg~~EmotDysreg = 0.479
EmotReact~~EmotReact = 0.447
BPD~~BPD = 0.488
DisordEat~~DisordEat = 0.778
Anxiety~~Anxiety = 0.608
Depression~1 = -0.089
Suicidality~1 = -0.115
EmotDysreg~1 = -0.083
EmotReact~1 = -0.073
BPD~1 = -0.104
DisordEat~1 = -0.063
Anxiety~1 = -0.06
N= 3244
3
N= 300 LR=64.6(df=22)
NSSI2 >= 5.5
f1=~Depression = 0.566
f1=~Suicidality = 0.674
f1=~EmotDysreg = 0.612
f1=~EmotReact = 0.615
f1=~BPD = 0.741
f1=~DisordEat = 0.239
f1=~Anxiety = 0.46
reg = -0.016
Depression~~Depression = 0.664
Suicidality~~Suicidality = 1.034
EmotDysreg~~EmotDysreg = 0.433
EmotReact~~EmotReact = 0.438
BPD~~BPD = 0.389
DisordEat~~DisordEat = 1.119
Anxiety~~Anxiety = 0.703
Depression~1 = 0.463
Suicidality~1 = 0.752
EmotDysreg~1 = 0.491
EmotReact~1 = 0.423
BPD~1 = 0.707
DisordEat~1 = 0.332
Anxiety~1 = 0.342
N= 163
f1=~Depression = 0.976
f1=~Suicidality = 0.799
f1=~EmotDysreg = 0.85
f1=~EmotReact = 0.863
f1=~BPD = 0.468
f1=~DisordEat = 0.583
f1=~Anxiety = 0.717
reg = 0.005
Depression~~Depression = 0.696
Suicidality~~Suicidality = 1.512
EmotDysreg~~EmotDysreg = 0.438
EmotReact~~EmotReact = 0.278
BPD~~BPD = 0.614
DisordEat~~DisordEat = 1.021
Anxiety~~Anxiety = 0.638
Depression~1 = 0.973
Suicidality~1 = 1.365
EmotDysreg~1 = 0.822
EmotReact~1 = 0.57
BPD~1 = 1.126
DisordEat~1 = 0.773
Anxiety~1 = 0.503
N= 137

Integrated Model Results

- BIC = 53417
  - Worse than the multiple group model with splits at 0.5
- A lot of parameters
  - What might improve the LRT will not necessarily improve fit when taking into account # of parameters
- Assumption: linear effects ***and*** model parameters differ between groups
- Would be best to only allow regression parameter to vary across groups
  - “focus.parameter=" in semtree

Covariates


1
N= 3512 LR=387.2(df=21)
no
NSSIFrequency >= 0.5
yes
2
N= 3088 LR=50.3(df=21)
11
N= 424 LR=54.1(df=21)
age >= 24.5
NSSIFrequency >= 5.5
INITIALIZED MODEL.A[1,8] = -0.681
INITIALIZED MODEL.A[2,8] = -0.5
INITIALIZED MODEL.A[3,8] = -0.609
INITIALIZED MODEL.A[4,8] = -0.72
INITIALIZED MODEL.A[5,8] = -0.693
INITIALIZED MODEL.A[6,8] = -0.336
INITIALIZED MODEL.A[7,8] = -0.753
INITIALIZED MODEL.S[1,1] = 0.42
INITIALIZED MODEL.S[2,2] = 0.581
INITIALIZED MODEL.S[3,3] = 0.524
INITIALIZED MODEL.S[4,4] = 0.442
INITIALIZED MODEL.S[5,5] = 0.509
INITIALIZED MODEL.S[6,6] = 0.631
INITIALIZED MODEL.S[7,7] = 0.47
INITIALIZED MODEL.M[1,1] = -0.082
INITIALIZED MODEL.M[1,2] = -0.003
INITIALIZED MODEL.M[1,3] = -0.389
INITIALIZED MODEL.M[1,4] = -0.184
INITIALIZED MODEL.M[1,5] = -0.032
INITIALIZED MODEL.M[1,6] = -0.174
INITIALIZED MODEL.M[1,7] = -0.212
N= 240
INITIALIZED MODEL.A[1,8] = -0.752
INITIALIZED MODEL.A[2,8] = -0.715
INITIALIZED MODEL.A[3,8] = -0.637
INITIALIZED MODEL.A[4,8] = -0.739
INITIALIZED MODEL.A[5,8] = -0.742
INITIALIZED MODEL.A[6,8] = -0.327
INITIALIZED MODEL.A[7,8] = -0.598
INITIALIZED MODEL.S[1,1] = 0.565
INITIALIZED MODEL.S[2,2] = 0.993
INITIALIZED MODEL.S[3,3] = 0.455
INITIALIZED MODEL.S[4,4] = 0.434
INITIALIZED MODEL.S[5,5] = 0.444
INITIALIZED MODEL.S[6,6] = 1.044
INITIALIZED MODEL.S[7,7] = 0.621
INITIALIZED MODEL.M[1,1] = 0.423
INITIALIZED MODEL.M[1,2] = 0.631
INITIALIZED MODEL.M[1,3] = 0.396
INITIALIZED MODEL.M[1,4] = 0.385
INITIALIZED MODEL.M[1,5] = 0.581
INITIALIZED MODEL.M[1,6] = 0.262
INITIALIZED MODEL.M[1,7] = 0.3
N= 287
INITIALIZED MODEL.A[1,8] = -0.985
INITIALIZED MODEL.A[2,8] = -0.8
INITIALIZED MODEL.A[3,8] = -0.852
INITIALIZED MODEL.A[4,8] = -0.863
INITIALIZED MODEL.A[5,8] = -0.47
INITIALIZED MODEL.A[6,8] = -0.581
INITIALIZED MODEL.A[7,8] = -0.722
INITIALIZED MODEL.S[1,1] = 0.693
INITIALIZED MODEL.S[2,2] = 1.513
INITIALIZED MODEL.S[3,3] = 0.44
INITIALIZED MODEL.S[4,4] = 0.281
INITIALIZED MODEL.S[5,5] = 0.614
INITIALIZED MODEL.S[6,6] = 1.021
INITIALIZED MODEL.S[7,7] = 0.634
INITIALIZED MODEL.M[1,1] = 1.094
INITIALIZED MODEL.M[1,2] = 1.459
INITIALIZED MODEL.M[1,3] = 0.925
INITIALIZED MODEL.M[1,4] = 0.672
INITIALIZED MODEL.M[1,5] = 1.181
INITIALIZED MODEL.M[1,6] = 0.839
INITIALIZED MODEL.M[1,7] = 0.588
N= 137
3
N= 2838 LR=48.1(df=21)
age >= 18.5
4
N= 406 LR=56.2(df=21)
7
N= 2432 LR=40.5(df=21)
gender in [Female]
age >= 20.5
INITIALIZED MODEL.A[1,8] = -0.759
INITIALIZED MODEL.A[2,8] = -0.321
INITIALIZED MODEL.A[3,8] = -0.715
INITIALIZED MODEL.A[4,8] = -0.592
INITIALIZED MODEL.A[5,8] = -0.577
INITIALIZED MODEL.A[6,8] = -0.627
INITIALIZED MODEL.A[7,8] = -0.828
INITIALIZED MODEL.S[1,1] = 0.377
INITIALIZED MODEL.S[2,2] = 0.247
INITIALIZED MODEL.S[3,3] = 0.433
INITIALIZED MODEL.S[4,4] = 0.386
INITIALIZED MODEL.S[5,5] = 0.518
INITIALIZED MODEL.S[6,6] = 0.474
INITIALIZED MODEL.S[7,7] = 0.275
INITIALIZED MODEL.M[1,1] = -0.243
INITIALIZED MODEL.M[1,2] = -0.243
INITIALIZED MODEL.M[1,3] = -0.051
INITIALIZED MODEL.M[1,4] = -0.233
INITIALIZED MODEL.M[1,5] = -0.147
INITIALIZED MODEL.M[1,6] = -0.243
INITIALIZED MODEL.M[1,7] = -0.388
N= 93
INITIALIZED MODEL.A[1,8] = -0.64
INITIALIZED MODEL.A[2,8] = -0.413
INITIALIZED MODEL.A[3,8] = -0.744
INITIALIZED MODEL.A[4,8] = -0.702
INITIALIZED MODEL.A[5,8] = -0.746
INITIALIZED MODEL.A[6,8] = -0.375
INITIALIZED MODEL.A[7,8] = -0.499
INITIALIZED MODEL.S[1,1] = 0.37
INITIALIZED MODEL.S[2,2] = 0.634
INITIALIZED MODEL.S[3,3] = 0.429
INITIALIZED MODEL.S[4,4] = 0.497
INITIALIZED MODEL.S[5,5] = 0.385
INITIALIZED MODEL.S[6,6] = 0.801
INITIALIZED MODEL.S[7,7] = 0.866
INITIALIZED MODEL.M[1,1] = 0.078
INITIALIZED MODEL.M[1,2] = 0.01
INITIALIZED MODEL.M[1,3] = 0.043
INITIALIZED MODEL.M[1,4] = 0.165
INITIALIZED MODEL.M[1,5] = 0.098
INITIALIZED MODEL.M[1,6] = 0.227
INITIALIZED MODEL.M[1,7] = 0.286
N= 312
INITIALIZED MODEL.A[1,8] = -0.618
INITIALIZED MODEL.A[2,8] = -0.387
INITIALIZED MODEL.A[3,8] = -0.664
INITIALIZED MODEL.A[4,8] = -0.684
INITIALIZED MODEL.A[5,8] = -0.626
INITIALIZED MODEL.A[6,8] = -0.411
INITIALIZED MODEL.A[7,8] = -0.566
INITIALIZED MODEL.S[1,1] = 0.417
INITIALIZED MODEL.S[2,2] = 0.538
INITIALIZED MODEL.S[3,3] = 0.492
INITIALIZED MODEL.S[4,4] = 0.456
INITIALIZED MODEL.S[5,5] = 0.466
INITIALIZED MODEL.S[6,6] = 0.849
INITIALIZED MODEL.S[7,7] = 0.635
INITIALIZED MODEL.M[1,1] = -0.123
INITIALIZED MODEL.M[1,2] = -0.151
INITIALIZED MODEL.M[1,3] = -0.053
INITIALIZED MODEL.M[1,4] = -0.087
INITIALIZED MODEL.M[1,5] = -0.146
INITIALIZED MODEL.M[1,6] = -0.036
INITIALIZED MODEL.M[1,7] = -0.079
N= 1396
INITIALIZED MODEL.A[1,8] = -0.634
INITIALIZED MODEL.A[2,8] = -0.404
INITIALIZED MODEL.A[3,8] = -0.724
INITIALIZED MODEL.A[4,8] = -0.754
INITIALIZED MODEL.A[5,8] = -0.632
INITIALIZED MODEL.A[6,8] = -0.366
INITIALIZED MODEL.A[7,8] = -0.566
INITIALIZED MODEL.S[1,1] = 0.579
INITIALIZED MODEL.S[2,2] = 0.536
INITIALIZED MODEL.S[3,3] = 0.409
INITIALIZED MODEL.S[4,4] = 0.411
INITIALIZED MODEL.S[5,5] = 0.52
INITIALIZED MODEL.S[6,6] = 0.704
INITIALIZED MODEL.S[7,7] = 0.541
INITIALIZED MODEL.M[1,1] = -0.08
INITIALIZED MODEL.M[1,2] = -0.168
INITIALIZED MODEL.M[1,3] = -0.081
INITIALIZED MODEL.M[1,4] = -0.081
INITIALIZED MODEL.M[1,5] = -0.124
INITIALIZED MODEL.M[1,6] = -0.134
INITIALIZED MODEL.M[1,7] = -0.054
N= 1036

Discussion

Limitations

- By not only focusing on invariance, not really sure how the groups differ
  - Also less straightforward with not estimating latent mean
- Relying solely on the BIC
  - Need to test if best non-nested information criterion
    - Tried and failed to come up with a good nested model comparison
- Computationally Intensive
  - Need to make the process easier
  - Not everyone wants to program 300 lines of code
  - Could be a new algorithm
    - Compare split improvement in LL to LL from MIMIC
      - Only split if cutoff improves
- Sample and variable specific
  - Best done with IDA
    - Can incorporate demographic effects

> 1 Variables

- Think of depression: Could have used items from a scale
- Dimensionality could increase quickly
- Maybe best to use lasso regularization to identify which variables contribute
  - Assumption that only a few variables differentiate individuals
  - SEM Trees implicitly does variable selection
  - Even better would be use variable importance from SEM Forests
    - Identify which variables are most important
      - Thus which stepwise effects and interactions

Thank You!

Email: rjacobuc@nd.edu for code or presentation

Questions/Comments?

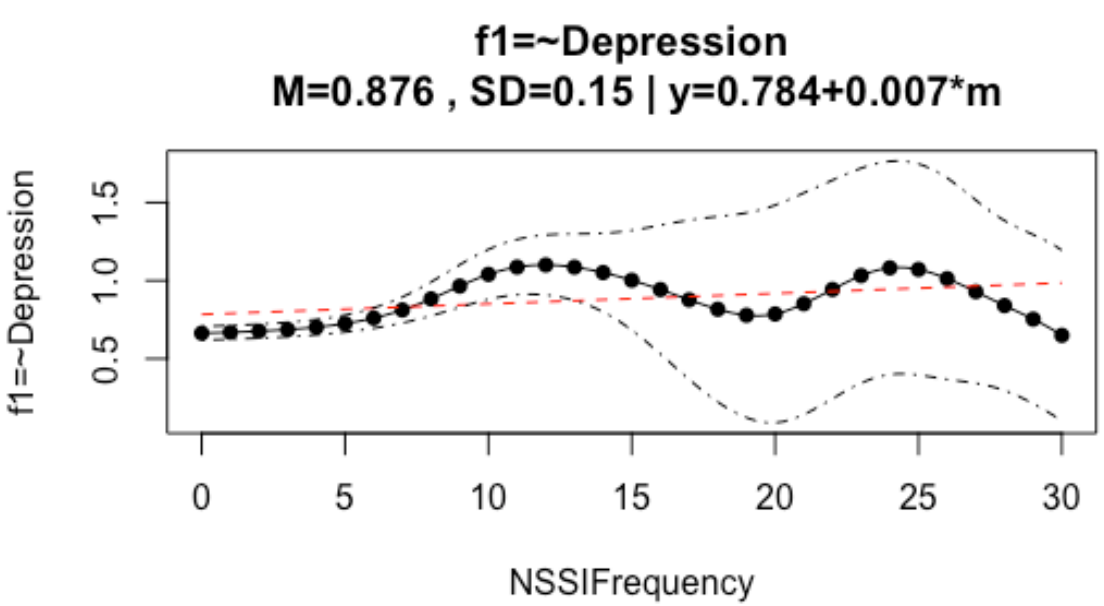
Purely Exploratory Model

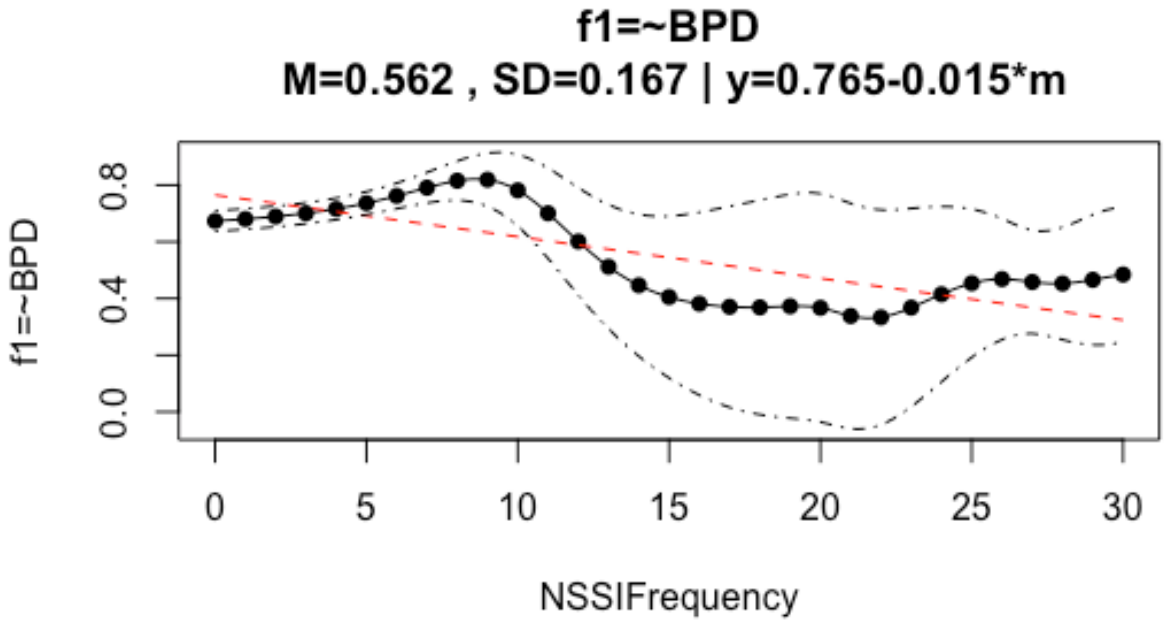
Loess SEM

- Uses Loess regression to assess how parameter estimates change across values of the moderator (NSSI)
  - Hildebrandt, A., Luedtke, O., Robitzsch, A., Sommer, C., & Wilhelm, O. (2016). Exploring factor model parameters across continuous variables with local structural equation models. *Multivariate Behavioral Research*, 51, 257-278.
  - Hildebrandt, A., Wilhelm, O., & Robitzsch, A. (2009). Complementary and competing factor analytic approaches for the investigation of measurement invariance. *Review of Psychology*, 16, 87-102.
- This is implemented in the sirt package:

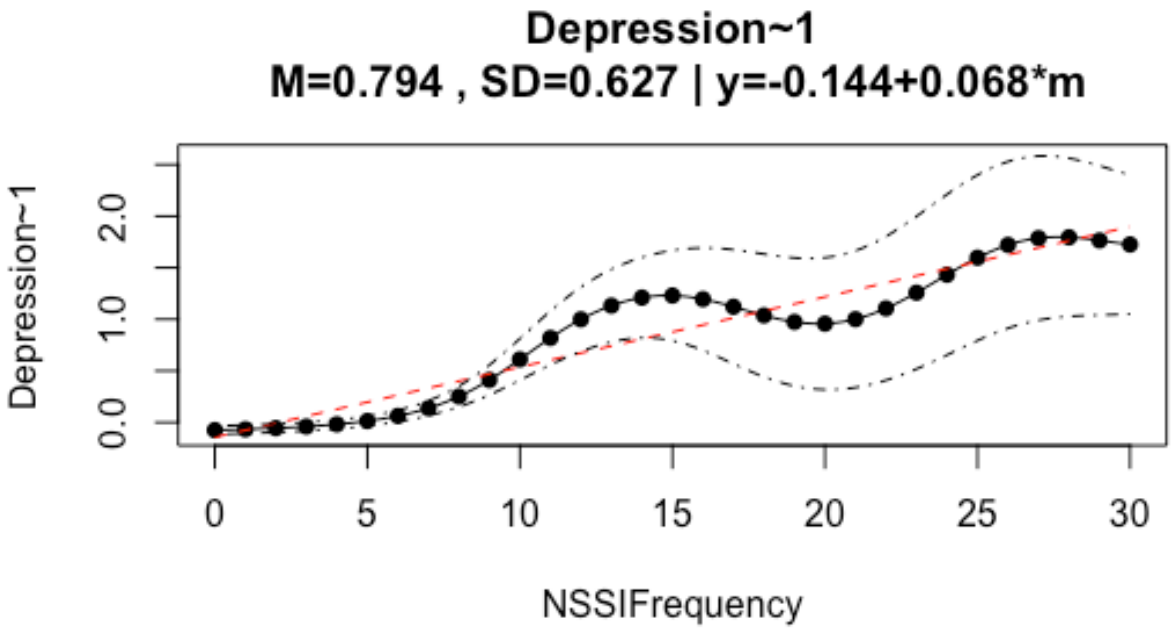
```
grid = 0:30
mod1 <- sirt::lsem.estimate(modelRev2.dat, moderator="NSSIFrequency",
                            moderator.grid=grid,h=.1,
                            lavmodel=cfa.mod11,residualize=F)

plot(mod1)
```

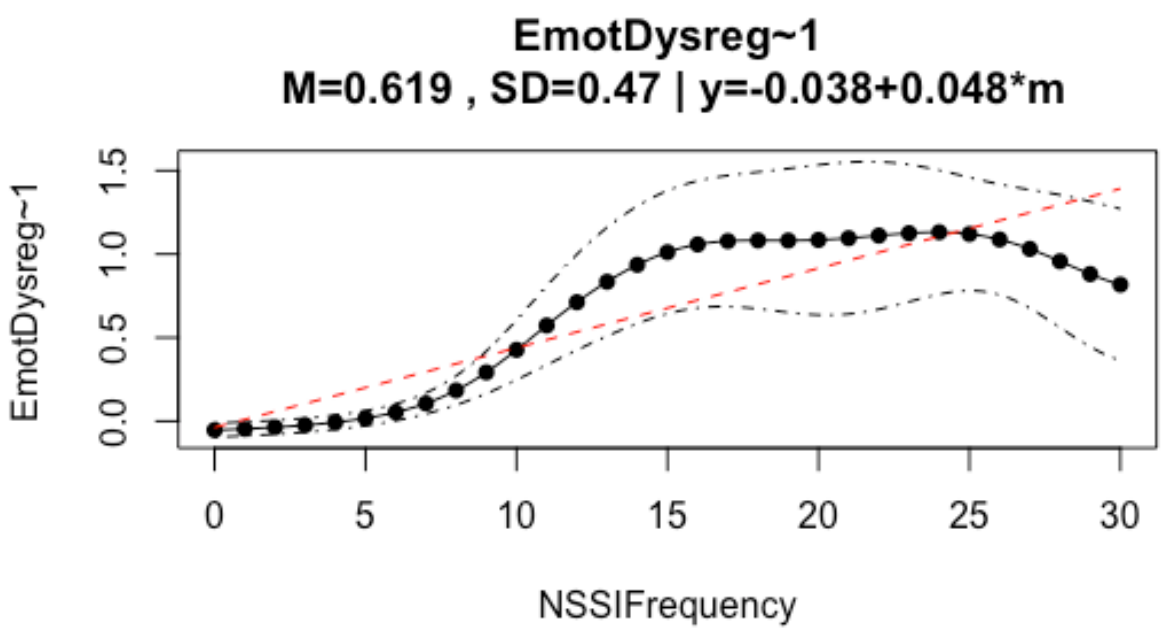
Factor Loadings





Mean Structure





Loess SEM: How can this help?

- This can alert us to measurement invariance violations.
- Maybe more importantly, how the means change
  - However, did not pick up small changes, as in between 0-1
  - But can give a more general picture
    - Smaller *spans* result in model estimation problems
- Can only use where you have enough data
  - Why I reduced NSSI to 0-30
- For a more principled analysis, need to translate partial ideas to models.

Loess SEM for Final Grouping