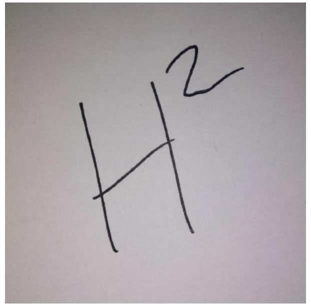
Assessing Model Parameter Stability in R

Hannah Frick

2020-02-29



Find me at...

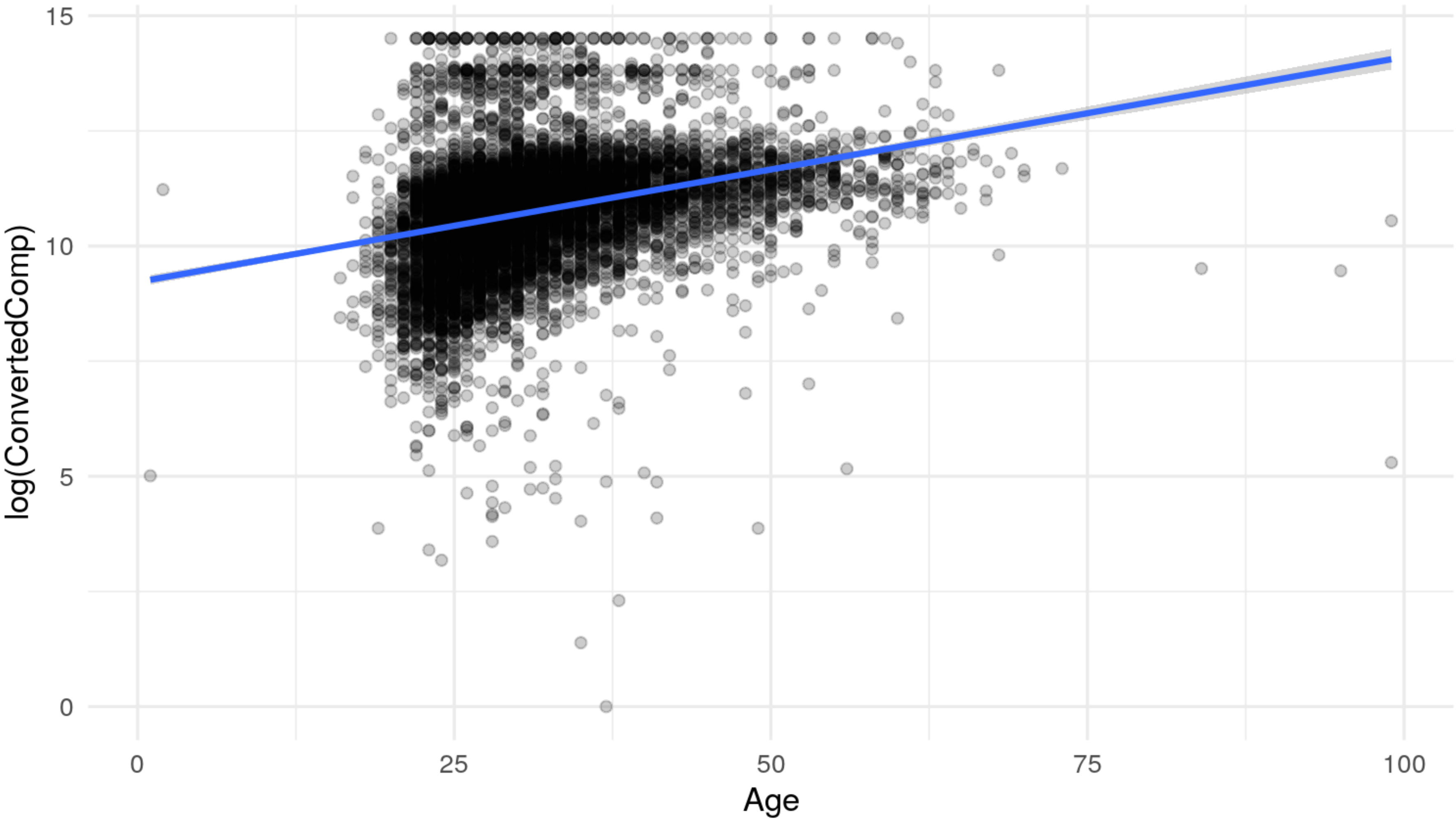
@hfcfrick
@hfrick
www.frick.ws

Let's talk about models!

StackOverflow developer survey 2019

- StackOverflow (SO): Q&A site for programming languages including R
- Yearly survey - In 2019: nearly 90k respondents of which 14k say they are data analysts/scientists, machine learning specialists, or academic researchers
- Questions on demographics, technology, work, and SO community
- Toy example here: Compensation ~ Age
- More insights from the survey by Julia Silge at https://insights.stackoverflow.com/survey/2019

Compensation ~ Age



Compensation ~ Age

```
so_lm <- lm(log(ConvertedComp) ~ Age, data = so)
summary(so_lm)
```

```
##
## Call:
## lm(formula = log(ConvertedComp) ~ Age, data = so)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.0272  -0.6809   0.0668   0.6339   4.3121
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 9.219313   0.055779  165.28   <2e-16 ***
## Age         0.048861   0.001694   28.85   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.291 on 8131 degrees of freedom
## Multiple R-squared:  0.09287,    Adjusted R-squared:  0.09276
## F-statistic: 832.4 on 1 and 8131 DF,  p-value: < 2.2e-16
```

Does this hold for the entire sample?

- You can include an interaction term. `log(ConvertedComp) ~ Age * JobSeek`
- You can include many interaction terms. `log(ConvertedComp) ~ Age * (JobSeek + LastHireDate + WorkPlan)`
- You can fit separate models and use a likelihood ratio test.
- But it's always you who picks which groups to look at.
- What if there are no other covariates to group by?
- Goal: find subgroups for which a set of model parameters holds - and if that's just one group we are in our usual case.
- Two approaches to establish groups in a data-driven way:
    - mixture models
    - model-based trees

Mixture Models

Mixture models

- Assumption: data stem from K different subgroups with unknown subgroup membership and subgroup-specific parameters.

- The full mixture model is a weighted sum over these separate models (or components):

$$f(y_i; x_i, \beta_{(1)}, \ldots, \beta_{(K)}) = \sum_{k=1}^{K} \pi_k \cdot f(y_i; x_i \beta_{(k)})$$

- If multiple components lead to better fit than a single model, you may have parameter instability.

- Use an information criterion, e.g. AIC or BIC, for that comparison.

- Estimation via the Expectation-Maximization (EM) algorithm: Alternate between

    - E-step: estimation of the posterior probabilities of each observation for the K components
    - M-step: estimation of the component models, weighted by the posterior probabilities

Mixture models

- In R: many different packages for specific types of mixture models.
- CRAN taskview https://cran.r-project.org/web/views/Cluster.html
- {flexmix}: provides the framework for the EM algorithm, you (can) provide the M-step.

```
library(flexmix)

so_mm <- stepFlexmix(log(ConvertedComp) ~ Age, data = so, k = 1:4, model = FLXMRglm(),
                     nrep = 5, control  = list(iter = 500))
```

```
## 1 : * * * * *
## 2 : * * * * *
## 3 : * * * * *
## 4 : * * * * *
```

```
BIC(so_mm)
```

```
##        1        2        3        4
## 27259.39 25940.51 25336.24 25321.86
```

Mixture models with the {flexmix} package

```r
so_mm_bic <- getModel(so_mm, which = "BIC")
summary(so_mm_bic)
```

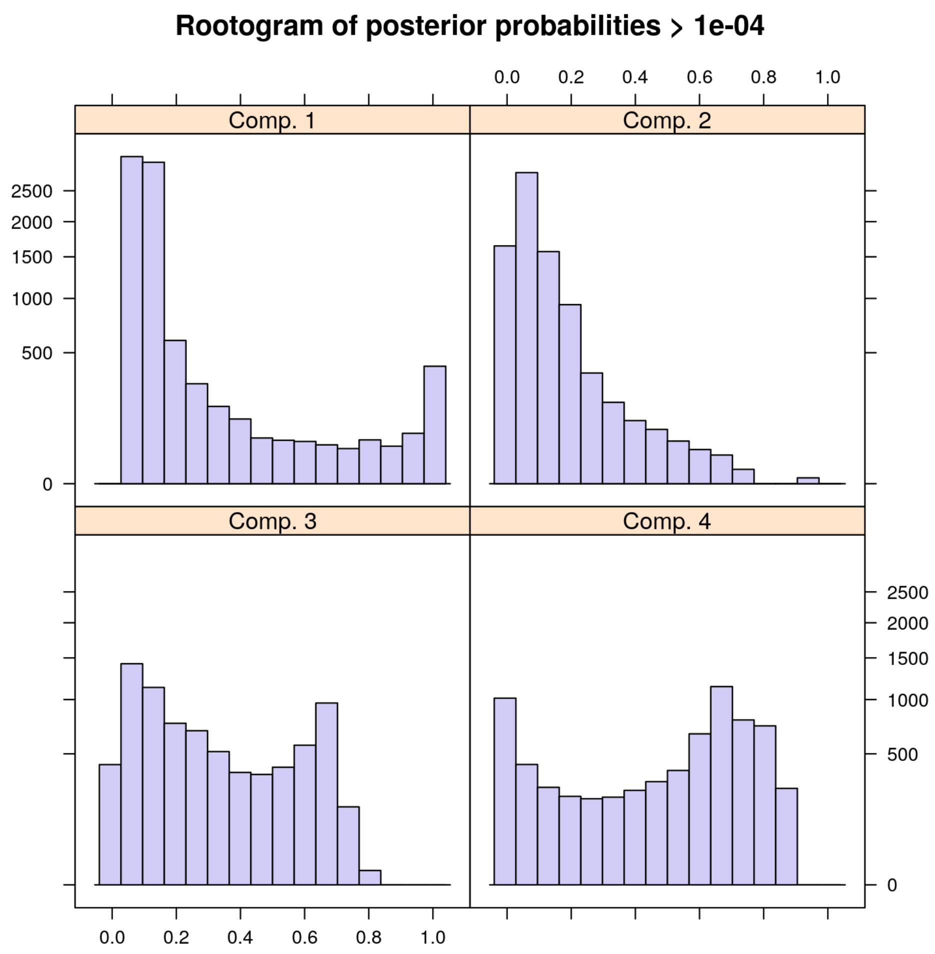
```
##
## Call:
## stepFlexmix(log(ConvertedComp) ~ Age, data = so, model = FLXMRglm(),
##     control = list(iter = 500), k = 4, nrep = 5)
##
##        prior size post>0  ratio
## Comp.1 0.197  828   8133 0.1018
## Comp.2 0.108  253   7839 0.0323
## Comp.3 0.293 2631   7789 0.3378
## Comp.4 0.402 4421   6941 0.6369
##
## 'log Lik.' -12593.4 (df=15)
## AIC: 25216.81   BIC: 25321.86
```

```r
parameters(so_mm_bic)
```

```
##                       Comp.1    Comp.2     Comp.3     Comp.4
## coef.(Intercept) 10.61270500 4.7104074 8.14855031 10.4961004
## coef.Age          0.01717237 0.1825487 0.06383459  0.0209600
## sigma             2.24366264 0.9303033 0.71742867  0.4336781
```

Mixture models with the {flexmix} package

```
plot(so_mm_bic)
```



Concomitant variables

- So far: *latent* groups - any covariates are part of the component models, *not* part of which observation belongs in which subgroup.
- Change that by using a model for the posterior probabilites, e.g., a multinomial logit model.
- Instead of non-parametric prior weights π_k we have $\pi(k|z, \alpha)$ which include the additional concomitant variables z .

```
so_mmc <- stepFlexmix(log(ConvertedComp) ~ Age, data = so, k = 1:4, model = FLXMRglm(),
                      concomitant = FLXPmultinom(~ WorkLoc + JobSeek + LastHireDate),
                      nrep = 5, control  = list(iter = 500))
```

```
## 1 : * * * * *
## 2 : * * * * *
## 3 : * * * * *
## 4 : * * * * *
```

```
BIC(so_mmc)
```

```
##        1        2        3        4 
## 27259.39 25778.11 25162.02 25163.38
```

```
so_mmc_bic <- getModel(so_mmc, which = "BIC")
```

Concomitant variables

```
c(BIC(so_mm_bic), BIC(so_mmc_bic))
```

```
## [1] 25321.86 25162.02
```

```
summary(so_mmc_bic)
```

```
##
## Call:
## stepFlexmix(log(ConvertedComp) ~ Age, data = so, model = FLXMRglm(),
##     concomitant = FLXPmultinom(~WorkLoc + JobSeek + LastHireDate),
##     control = list(iter = 500), k = 3, nrep = 5)
##
##        prior size post>0 ratio
## Comp.1 0.301 2372   7828 0.303
## Comp.2 0.219  898   8133 0.110
## Comp.3 0.480 4863   7255 0.670
##
## 'log Lik.' -12459.46 (df=27)
## AIC: 24972.92   BIC: 25162.02
```

Concomitant variables

```r
parameters(so_mmc_bic)
```

```
##                    Comp.1      Comp.2      Comp.3
## coef.(Intercept) 7.3559217 10.04003820 10.36734458
## coef.Age         0.0846721  0.03554901  0.02312614
## sigma            0.7681744  2.20440021  0.49064397
```

```r
parameters(so_mmc_bic, which = "concomitant")
```

```
##                  1           2           3
## (Intercept)      0  0.65288825   1.5215181
## WorkLocOffice    0 -0.06786797  -0.1080250
## WorkLocOther     0 -0.31743198  -0.3115621
## JobSeekOpen      0 -1.24094144  -1.5593537
## JobSeekYes       0 -1.52506594  -2.2986761
## LastHireDate1-2y 0  0.04868279   0.2509607
## LastHireDate3-4y 0  0.53519814   0.8712738
## LastHireDate4+y  0  0.48900486   0.8875694
## LastHireDateNA   0 -1.04920539 -11.9820206
```

Concomitant variables

```
so_mmc_bic_rf <- refit(so_mmc_bic)
summary(so_mmc_bic_rf, which = "concomitant")
```

```
## $Comp.2
##                    Estimate Std. Error z value  Pr(>|z|)
## (Intercept)        0.708524   0.241720  2.9312  0.003377 **
## WorkLocOffice     -0.064710   0.125300 -0.5164  0.605546
## WorkLocOther      -0.319736   0.191422 -1.6703  0.094857 .
## JobSeekOpen       -1.265698   0.198954 -6.3618 1.995e-10 ***
## JobSeekYes        -1.557330   0.242676 -6.4173 1.387e-10 ***
## LastHireDate1-2y   0.050277   0.138537  0.3629  0.716670
## LastHireDate3-4y   0.539672   0.180834  2.9843  0.002842 **
## LastHireDate4+y    0.492031   0.181354  2.7131  0.006666 **
## LastHireDateNA    -1.047742   0.852925 -1.2284  0.219293
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## $Comp.3
##                    Estimate Std. Error  z value  Pr(>|z|)
## (Intercept)         1.57720    0.20377   7.7402 9.926e-15 ***
## WorkLocOffice      -0.10173    0.10012  -1.0161   0.30960
## WorkLocOther       -0.30804    0.14766  -2.0861   0.03697 *
## JobSeekOpen        -1.57276    0.16306  -9.6452 < 2.2e-16 ***
## JobSeekYes         -2.31633    0.19679 -11.7708 < 2.2e-16 ***
## LastHireDate1-2y    0.24916    0.10530   2.3662   0.01797 *
## LastHireDate3-4y    0.87818    0.14242   6.1663 6.990e-10 ***
## LastHireDate4+y     0.88445    0.15233   5.8061 6.395e-09 ***
## LastHireDateNA    -11.98204  133.11955  -0.0900   0.92828
```

Model-Based Trees

Model-based trees

Generate a tree as follows (Zeileis et al, 2008)

1) Estimate the model parameters in the current subgroup.

2) Test parameter stability along each partitioning variable (Zeileis et al, 2007).

3) If any instability is found, split the sample along the variable with the highest instability.
Choose the breakpoint with the highest improvement in model fit.

4) Repeat 1--3 on the resulting subsamples until no further instability is found.

[1] Zeileis A, Hothorn T, Hornik K (2008). “Model-Based Recursive Partitioning.” Journal of Computational and Graphical Statistics, 17(2), 492-514. doi:10.1198/106186008X319331

[2] Zeileis A, Hornik K (2007). "Generalized M-Fluctuation Tests for Parameter Instability." Statistica Neerlandica, 61(4), 488-508. doi:10.1111/j.1467-9574.2007.00371.x

Model-based trees with the {partykit} package

```
library(partykit)

so_tree <- lmtree(log(ConvertedComp) ~ Age | WorkLoc + JobSeek + LastHireDate, data = so)
plot(so_tree, terminal_panel = NULL)
```

Comparison

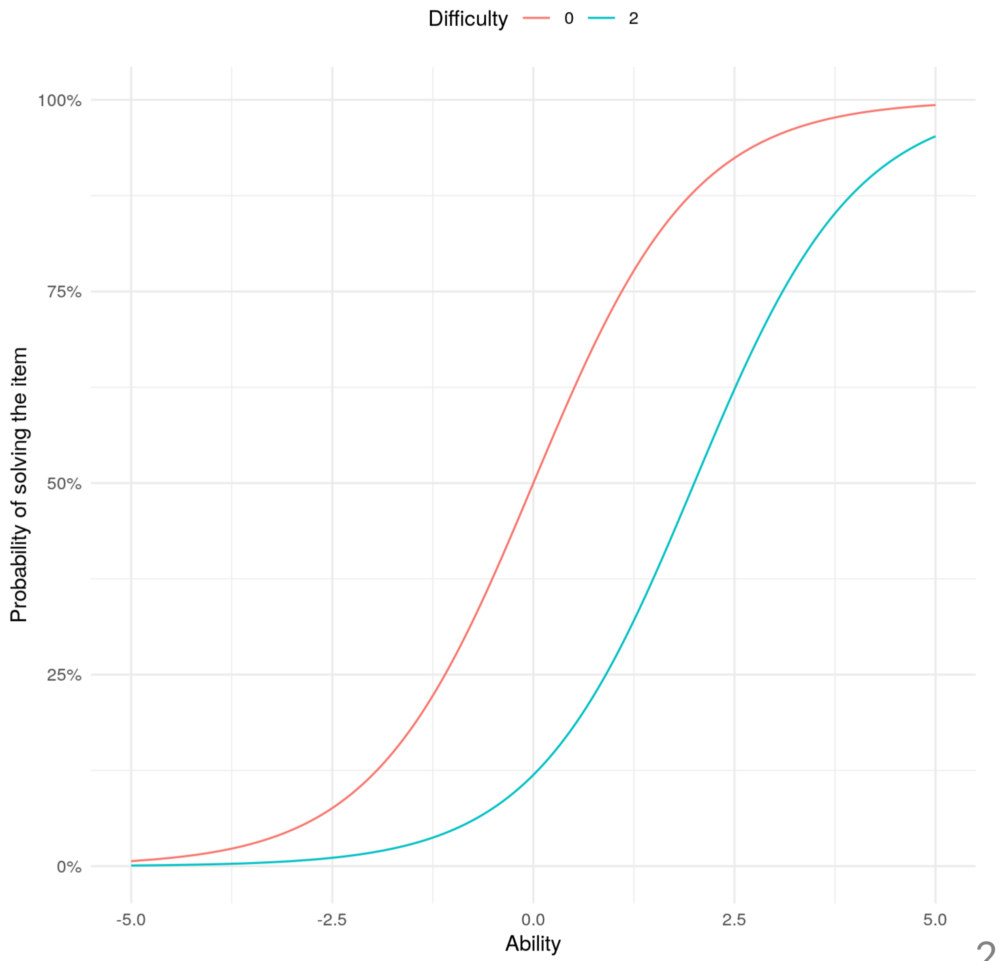
Comparison of mixture models and model-based trees

- *Selection of number of subgroups*: via infomation criterion for mixture model, via significance tests for trees.
- *Covariates*: optional for mixture models, required for trees.
- *Link between covariates and subgroups*: smooth, monotonic transition between subgroups for mixture models, abrupt shifts for trees which make non-monotonic transitions possible (through multiple shifts).
- *Variable selection on spliting variables*: requires an additional step for mixture models, is inherent for trees.
- *Clustering*: mixture models yield a probabilistic clustering, trees yield a hard clustering.

Application in Psychometrics

Rasch model for latent traits

- Probabilistic model for measuring latent traits.
- Subject repond to several (binary) items.
- Probability of solving an item depends (only!) on the item difficulty and the subject ability.
- Central assumption:
  measurement invariance == parameter stability!



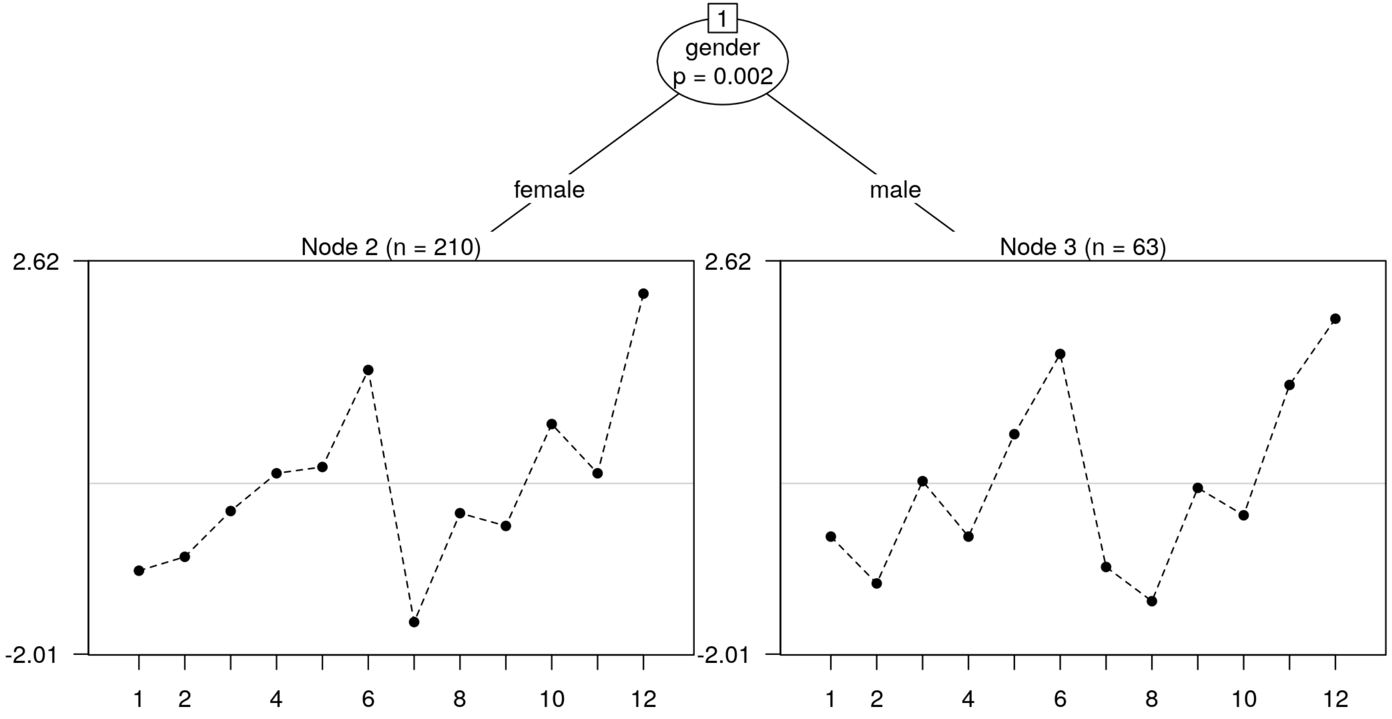
Verbal aggression

- 12 items which combine
    - A frustrating situation:
        - *A bus fails to stop for me.*
        - *I miss a train because a clerk gave me faulty information.*
    - A behavioural mode:
        - *want*
        - *do*
    - A response:
        - *curse*
        - *scold*
        - *shout*
- For example: "A bus fails to stop for me. I want to curse."
- 316 subjects
- 2 covariates: gender and an anger score

Rasch tree with the {psychotree} package

```r
library(psychotree)

va_tree <- raschtree(resp2 ~ gender + anger, data = verbal_aggression)
plot(va_tree)
```



Rasch mixture model with the {psychomix} package

```r
library(psychomix)

va_mm <- raschmix(resp2 ~ 1, data = verbal_aggression,
                  k = 1:4,  scores = "meanvar", restricted = TRUE, nrep = 5)
```

```
## 1 : * * * * *
## 2 : * * * * *
## 3 : * * * * *
## 4 : * * * * *
```

```r
va_mm_c <- raschmix(resp2 ~ gender + anger, data = verbal_aggression,
                    k = 1:4, scores = "meanvar", restricted = TRUE, nrep = 5)
```

```
## 1 : * * * * *
## 2 : * * * * *
## 3 : * * * * *
## 4 : * * * * *
```

Frick H, Strobl C, Zeileis A (2015). Rasch Mixture Models for DIF Detection: A Comparison of Old and New Score Specifications. Educational and Psychological Measurement, 75(2).

Rasch mixture model with the {psychomix} package

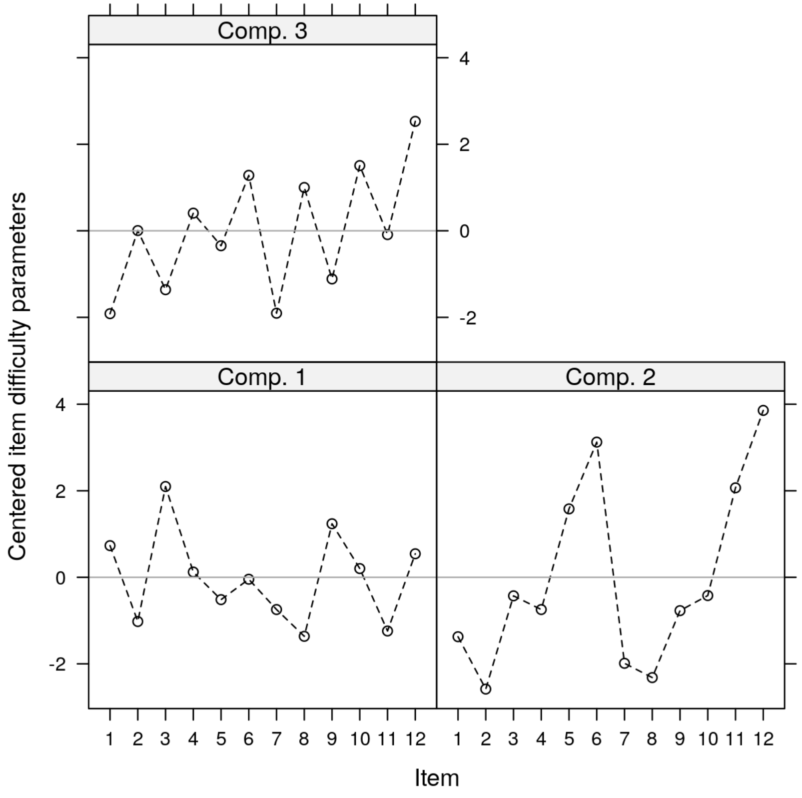
```r
rbind(mix = BIC(va_mm), mixC = BIC(va_mm_c))
```

```
##             1        2        3        4
## mix  3874.632 3847.766 3841.363 3858.837
## mixC 3874.632 3854.706 3843.681 3860.493
```

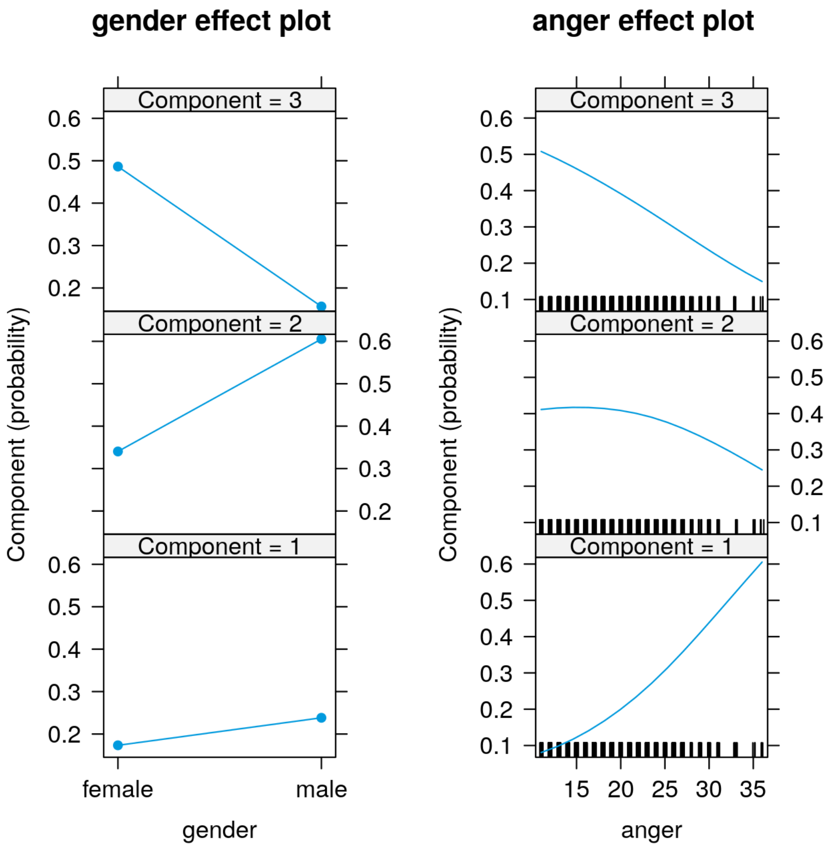
```r
va_mix <- getModel(va_mm_c, which = "3")
```

Rasch mixture model with the {psychomix} package

```
xyplot(va_mix)
```




```
effectsplot(va_mix)
```



Summary

Summary

- One set of model parameters may not always hold for the entire sample.
- Mixture models and model-based trees are two approaches to check for parameter instability in a data-driven way.
- Both come with their own strengths and weaknesses, try out both in practice.
- Implementations in {flexmix} and {partykit} are extensible to further models.
- Examples for such extensions in {psychomix} and {psychotree}.

Thanks

- Achim Zeileis, Carolin Strobl, and Bettina Grün for shared work
- Maëlle Salmon for feeback on this talk
- Alison Hill for the R-Ladies xaringan theme