Latent Variable Effect Forests

Gesis Workshop Contribution

Franz Classe [1] Christoph Kern [2]

[1]Bavarian State Institute for Higher Education Research and Planning

[2]University of Mannheim

October 14, 2021

Latent Variable Modeling

Definition:

Latent Variables are **unobservable phenomena** like

- creativity (Jauk et al., 2014)
- extraversion, neuroticism
- social anxiety, depression (Prenoveau et al., 2011)
- psychopathic personality (Drislane & Patrick, 2017)
- self-leadership (Furtner et al., 2015)

measured as theoretical **constructs** through research tools like a questionnaire in a survey.

Latent Variable Modeling

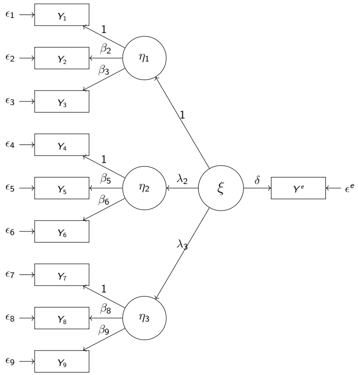
Definition:

Latent Variables are **unobservable phenomena** like

- creativity (Jauk et al., 2014)
- extraversion, neuroticism
- social anxiety, depression (Prenoveau et al., 2011)
- psychopathic personality (Drislane & Patrick, 2017)
- self-leadership (Furtner et al., 2015)

measured as theoretical **constructs** through research tools like a questionnaire in a survey.

Example:



Validity

Validity is...

> *"...the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests."*
>
> *(APA, 2014)*

Need for researchers to find **evidence to support proposed interpretation** of item responses.

Four sources of evidence for construct validity (APA, 2014):

- Test content

Validity

Validity is...

> *"...the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests."*
>
> *(APA, 2014)*

Need for researchers to find **evidence to support proposed interpretation** of item responses.

<u>Four sources of evidence</u> for construct validity (APA, 2014):

- Test content
- Response processes

Validity

Validity is...

> *"...the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests."*
>
> *(APA, 2014)*

Need for researchers to find **evidence to support proposed interpretation** of item responses.

<u>Four sources of evidence</u> for construct validity (APA, 2014):

- Test content
- Response processes
  $\rightarrow$ differential item functioning

Validity

Validity is...

> *"...the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests."*
>
> *(APA, 2014)*

Need for researchers to find **evidence to support proposed interpretation** of item responses.

Four sources of evidence for construct validity (APA, 2014):

- Test content
- Response processes
  $\rightarrow$ differential item functioning
- Internal structure

Validity

Validity is...

> *"...the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests."*
>
> *(APA, 2014)*

Need for researchers to find **evidence to support proposed interpretation** of item responses.

<u>Four sources of evidence</u> for construct validity (APA, 2014):

- Test content
- Response processes
  $\rightarrow$ differential item functioning
- Internal structure
  $\rightarrow$ model fit heterogeneity

Validity

Validity is...

> *"...the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests."*
>
> *(APA, 2014)*

Need for researchers to find **evidence to support proposed interpretation** of item responses.

Four sources of evidence for construct validity (APA, 2014):

- Test content
- Response processes
  $\rightarrow$ differential item functioning
- Internal structure
  $\rightarrow$ model fit heterogeneity
- Relation to other variables

Validity

Validity is...

> *"...the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests."*
>
> *(APA, 2014)*

Need for researchers to find **evidence to support proposed interpretation** of item responses.

Four sources of evidence for construct validity (APA, 2014):

- Test content
- Response processes
  $\rightarrow$ differential item functioning
- Internal structure
  $\rightarrow$ model fit heterogeneity
- Relation to other variables
  $\rightarrow$ effect heterogeneity

Goals

How can we use latent variable modeling together with **machine learning techniques** to **detect heterogeneity** in the sample and to **estimate valid latent variable scores**?

Goals

How can we use latent variable modeling together with **machine learning techniques** to **detect heterogeneity** in the sample and to **estimate valid latent variable scores**?

1. Choose latent variable modeling framework that allows computationally efficient parameter estimation

Goals

How can we use latent variable modeling together with **machine learning techniques** to **detect heterogeneity** in the sample and to **estimate valid latent variable scores**?

1. Choose latent variable modeling framework that allows computationally efficient parameter estimation
2. Create method that automatically searches for subgroups in which for every group member
    1. items work in the same way

Goals

How can we use latent variable modeling together with **machine learning techniques** to **detect heterogeneity** in the sample and to **estimate valid latent variable scores**?

1. Choose latent variable modeling framework that allows computationally efficient parameter estimation
2. Create method that automatically searches for subgroups in which for every group member
    1. items work in the same way
    2. internal structure of questionnaire applies

Goals

How can we use latent variable modeling together with **machine learning techniques** to **detect heterogeneity** in the sample and to **estimate valid latent variable scores**?

1. Choose latent variable modeling framework that allows computationally efficient parameter estimation
2. Create method that automatically searches for subgroups in which for every group member
    1. items work in the same way
    2. internal structure of questionnaire applies
    3. relation to other variable is the same

Goals

How can we use latent variable modeling together with **machine learning techniques** to **detect heterogeneity** in the sample and to **estimate valid latent variable scores**?

1. Choose latent variable modeling framework that allows computationally efficient parameter estimation
2. Create method that automatically searches for subgroups in which for every group member
    1. items work in the same way
    2. internal structure of questionnaire applies
    3. relation to other variable is the same
3. Estimate latent variable scores for subgroups that fulfill requirements

LVE Forest

Model based recursive partitioning: Algorithm to recursively partition data set to reduce parameter instability (Zeileis et al., 2008)

1. Partition data set to control for DIF and effect heterogeneity (*tree*)

LVE Forest

Model based recursive partitioning: Algorithm to recursively partition data set to reduce parameter instability (Zeileis et al., 2008)

1. Partition data set to control for DIF and effect heterogeneity (*tree*)
   - Draw random subsample to grow tree & estimate model with rest of the sample (*double sampling*)(Athey & Imbens, 2016)

LVE Forest

Model based recursive partitioning: Algorithm to recursively partition data set to reduce parameter instability (Zeileis et al., 2008)

1. Partition data set to control for DIF and effect heterogeneity (*tree*)
   - Draw random subsample to grow tree & estimate model with rest of the sample (*double sampling*)(Athey & Imbens, 2016)
2. Repeat process with variation at every iteration (*forest*)

LVE Forest

Model based recursive partitioning: Algorithm to recursively partition data set to reduce parameter instability (Zeileis et al., 2008)

1. Partition data set to control for DIF and effect heterogeneity (*tree*)
   - Draw random subsample to grow tree & estimate model with rest of the sample (*double sampling*)(Athey & Imbens, 2016)
2. Repeat process with variation at every iteration (*forest*)
   - Randomly sample partitioning variables (*random split selection*) (Breiman, 2001)

LVE Forest

Model based recursive partitioning: Algorithm to recursively partition data set to reduce parameter instability (Zeileis et al., 2008)

1. Partition data set to control for DIF and effect heterogeneity (*tree*)
   - Draw random subsample to grow tree & estimate model with rest of the sample (*double sampling*)(Athey & Imbens, 2016)
2. Repeat process with variation at every iteration (*forest*)
   - Randomly sample partitioning variables (*random split selection*) (Breiman, 2001)
3. Select relevant subgroups based on
   - model fit (RMSEA-cutoff)

LVE Forest

Model based recursive partitioning: Algorithm to recursively partition data set to reduce parameter instability (Zeileis et al., 2008)

1. Partition data set to control for DIF and effect heterogeneity (*tree*)
   - Draw random subsample to grow tree & estimate model with rest of the sample (*double sampling*)(Athey & Imbens, 2016)
2. Repeat process with variation at every iteration (*forest*)
   - Randomly sample partitioning variables (*random split selection*) (Breiman, 2001)
3. Select relevant subgroups based on
   - model fit (RMSEA-cutoff)
   - direction of latent variable effect (effect parameter-cutoff)

LVE Forest

Model based recursive partitioning: Algorithm to recursively partition data set to reduce parameter instability (Zeileis et al., 2008)

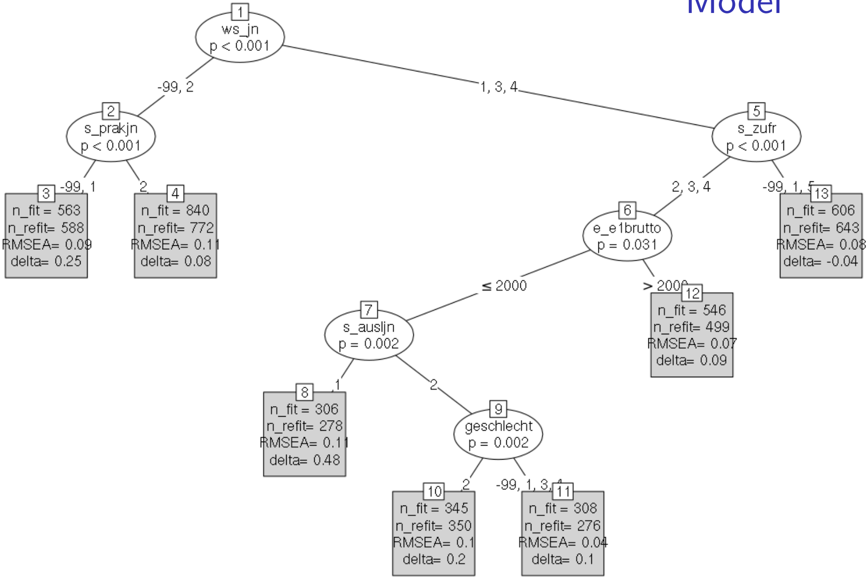
1. Partition data set to control for DIF and effect heterogeneity (*tree*)
   - Draw random subsample to grow tree & estimate model with rest of the sample (*double sampling*)(Athey & Imbens, 2016)
2. Repeat process with variation at every iteration (*forest*)
   - Randomly sample partitioning variables (*random split selection*) (Breiman, 2001)
3. Select relevant subgroups based on
   - model fit (RMSEA-cutoff)
   - direction of latent variable effect (effect parameter-cutoff)
4. Use relevant subgroups to predict latent variable scores

LVE Forest

Model based recursive partitioning: Algorithm to recursively partition data set to reduce parameter instability (Zeileis et al., 2008)

1. Partition data set to control for DIF and effect heterogeneity (*tree*)
   - Draw random subsample to grow tree & estimate model with rest of the sample (*double sampling*)(Athey & Imbens, 2016)
2. Repeat process with variation at every iteration (*forest*)
   - Randomly sample partitioning variables (*random split selection*) (Breiman, 2001)
3. Select relevant subgroups based on
   - model fit (RMSEA-cutoff)
   - direction of latent variable effect (effect parameter-cutoff)
4. Use relevant subgroups to predict latent variable scores
5. Compile results across all iterations

Tree in LVE Forest for Self-Leadership Model



Conclusion

LVE Forest detects and controls for:

- Heterogeneity of response styles/understanding (control for DIF)
- Heterogeneity of internal structure (RMSEA-cutoff)
- Heterogeneity of relation to other constructs (control for effect heterogeneity & effect parameter-cutoff)

Conclusion

LVE Forest detects and controls for:

- Heterogeneity of response styles/understanding (control for DIF)
- Heterogeneity of internal structure (RMSEA-cutoff)
- Heterogeneity of relation to other constructs (control for effect heterogeneity & effect parameter-cutoff)

$\rightarrow$ Predicted latent variable scores fulfill criteria for validity
although construct may not generally be valid

References

American Psychological Association. (2014). Standards for psychological and educational testing. *New York: American Educational Research Association*.

Athey, S., & Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, *113*(27), 7353–7360.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Drislane, L. E., & Patrick, C. J. (2017). Integrating alternative conceptions of psychopathic personality: A latent variable model of triarchic psychopathy constructs. *Journal of personality disorders*, *31*(1), 110–132.

Furtner, M. R., Rauthmann, J. F., & Sachse, P. (2015). Unique self-leadership: A bifactor model approach. *Leadership*, *11*(1), 105–125.

Jauk, E., Benedek, M., & Neubauer, A. C. (2014). The road to creative achievement: A latent variable model of ability and personality predictors. *European journal of personality*, *28*(1), 95–105.

Prenoveau, J. M., Craske, M. G., Zinbarg, R. E., Mineka, S., Rose, R. D., & Griffith, J. W. (2011). Are anxiety and depression just as stable as personality during late adolescence? results from a three-year longitudinal latent variable study.. *Journal of Abnormal Psychology*, *120*(4), 832.

Zeileis, A., Hothorn, T., & Hornik, K. (2008). Model-based recursive partitioning. *Journal of Computational and Graphical Statistics*, *17*(2), 492–514.

Any questions?

- Contact: classe@ihf.bayern.de
- Want to try the method?: `scforest` on github (coming soon)