

# Combining regression trees and linear regression to detect a specific type of interaction effects

Elise Dusseldorp and Jacqueline Meulman  
Data Theory Group, Leiden University, the Netherlands  
[dusseldorp@fsw.leidenuniv.nl](mailto:dusseldorp@fsw.leidenuniv.nl)

**Keywords:** Regression trees, Interaction

## Abstract

The regression trunk approach (RTA) combines two analysis methods: regression trees (Breiman, Friedman, Olshen & Stone, 1984) and multiple linear regression analysis. RTA detects interaction effects between predictors, in the regression of one continuous response variable on multiple predictor variables (see Dusseldorp & Meulman, 2001). In this paper, we compare the performances of RTA and forward stepwise linear regression in identifying significant treatment covariate interactions (TCIs), using simulation studies. A TCI is defined as the interaction between a treatment variable  $\mathbf{t}$  (dichotomous) and a covariate  $\mathbf{x}$  (continuous or categorical). The presence of a TCI indicates that the effect of  $\mathbf{t}$  differs for subjects with different values on  $\mathbf{x}$ . Common practice in regression analysis is to represent interactions as cross-products; in this case,  $\mathbf{t} \times \mathbf{x}$ . This representation implies that the effect of  $\mathbf{t}$  on a response variable differs over the whole range of  $\mathbf{x}$ . We focus on an alternative type of interaction, namely a threshold interaction, implying that  $\mathbf{t}$  has an effect only above or below a certain threshold value on  $\mathbf{x}$ . A regression tree is a convenient representation of this type of interaction. In two simulation studies we generate prediction models with one and two first-order threshold interactions. Three factors are included in the design: the ratio of the number of subjects and the number of covariates, the size of the TCI effect (from zero to large), and the group size (the number of subjects for whom the treatment is especially effective). In this way, we could also examine the power and the Type-I errors of RTA. The results show that RTA detects the interactions in a higher number of cases than forward stepwise regression. The power of RTA appears to be reasonable for medium to large TCI effect sizes, depending on the number of true TCIs in the model.

## References

- Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and regression trees*. Belmont, CA: Wadsworth.
- Dusseldorp, E. and Meulman, J. J. (2001). Prediction in medicine by integrating regression trees into regression analysis with optimal scaling. *Methods of Information in Medicine*, 40, 403-409.