Tutorial of moderated mediation with SEM: the gemm package

Peter Verboon[1] & Gjalt-Jorn Peters [1]

[1] Open University

Abstract

 Moderated mediation analysis assumes that the effect of a predictor on a dependent variable is mediated by one or more mediators and that the indirect paths in this mediation are moderated. Moderation can take place on the effect of the predictor on the mediators, or on the effect of the mediators on the dependent variable.

In this tutorial we will show how the package "gemm" can be used to do a moderated mediation analysis.

Keywords: moderated mediation SEM Lavaan

Tutorial of moderated mediation with SEM: the gemm package

Introduction

In modeling the relation between a predictor and a dependent variable there are basically three modeling situations to consider. First, there are other variables that are correlated with the dependent variable, and possibly also with the predictor. Adding such variables, which are usually called covariates, to the regression model may change the regression coefficient of the predictor, and the standard error for this estimate. This is discussed in many textbooks about multivariate and regression analysis. Second, the mechanism explaining the effect between predictor and dependent can be examined by a mediation analysis. The mediator is a variable that partly "explains" the assumed causal process, which links the predictor to the dependent variable. Third, the effect of interest depends on another variable, which is called the moderator. A moderator influences the effect of the predictor on the dependent, implying that the effect is conditional on the value of the moderator. When the three modeling situations are combined we obtain a general model, which is useful in many applied research. This general model is called the moderated mediation model and is earlier described in the methodological literature (e.g. Edwards and Lambert (2007); Hayes (2015); Preacher, Rucker, and Hayes (2007)). In this paper a short description of the moderated mediation model is given. Next the R package "gemm" is discussed, which is developed to analyze this model in an easy and straightforward way. A moderated mediation model assumes that the effect of a predictor on a dependent variable is mediated by one or more mediators and that the indirect paths in this mediation are moderated. Moderation can take place on the effect of the predictor on the mediators, or on the effect of the mediators on the dependent variable. In this tutorial the function `gemm()` is described for the analysis of moderated mediation models. The function is an alternative for the (SPSS or SAS) PROCESS macro developed by Hayes (Hayes, 2018). The `gemm` is in the R package `gemm`. This package can be installed

from Github and then loaded using the `library()` function:

    devtools::install_github("PeterVerboon/gemm")

```
library(gemm)
```

An illustration of the moderated mediation model, for which the function can be applied, is shown in Figure 1. The figure represents a structural model, linking the observed variables to each other. We are therefore using the SEM software to fit the moderated mediation model. To this end we use the lavaan package (Rosseel, 2012). Figure 1 shows the most general situation, with several mediators, several covariates for the mediators and several covariates for the dependent variable. There are also two moderators in this model, one for the paths from the predictor to the mediators and one for the paths from the mediators to the dependent variabele.

The model is limited in the sense that only one predictor, one dependent variable, and for each path one moderator are defined. Furthermore, serial mediation is not included in this model. The number of mediators and covariates is not limited.

The function `gemm()` calls the functions `buildModMedSemModel()`, which builds the requested SEM model. In this model we assume that the covariates for the mediators (C) are correlated with each other, and furthermore that the covariates fir the dependent variable (D) correlate with each other and also with the mediators. If covariates (C) are included in the model, paths are assumed to exist from the covariates (C) to all mediators. For the function to run the model needs to have a predictor, a dependent variable and at least one mediator. All other variables are optional. This most basic model (predictor –> mediator –> dependent) is just identified: three regression coefficients are estimated from the covariance matrix, containing three correlations. With two mediators there are six datapoints, and als six parameters to estimate (5 regression paths and one covariance between the mediators). With three mediators there are ten datapoints, and also ten
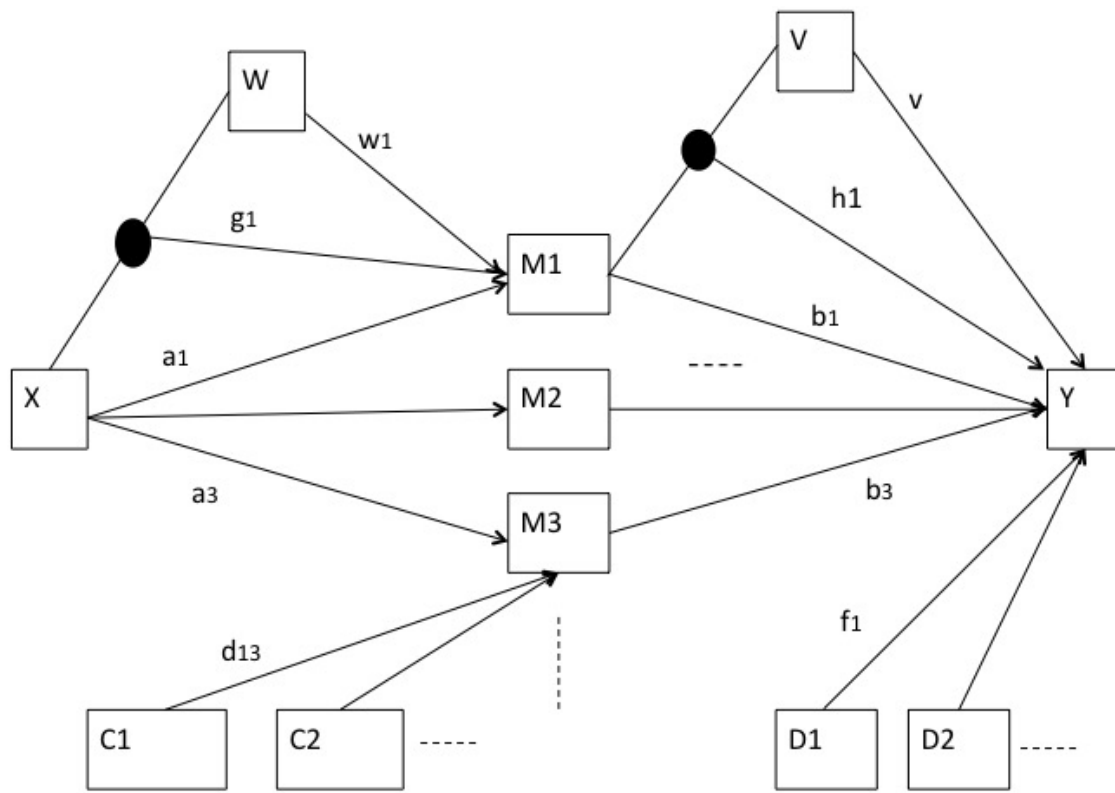
Figure 1. Moderated mediation model

parameters to estimate (three regression paths and the three covariances between the mediators). It is easy to see that models with only mediators, a predictor and a dependent variable, are always just identified. The fit values for these model s will therefore show perfect fit. Adding covariates or moderators will result in models in which the parameters have to be estimated iteratively, yielding model fit values that indicate how well the model fits the data. After running the function `gemm()` there are several functions available that are designed to handle the results: `plotSS()`, `plotIMM()`, `plotIMM3d()` and `print()`. These functions will be illustrated in this tutorial.

Index of moderated mediation

A primary statistic of interest in moderated mediation models is the index moderated mediation (Hays, 2015), which can be plotted by `gemm`. For a good understanding of this index, we will first present its algebraic derivation. When there are no moderators or when moderators take the value zero, the indirect (unmoderated) effect is defined as: $a_j b_j$. The indirect moderated effect with the X – M path moderated by W (assuming there is no moderation through V or when V = 0) is derived as follows:

$$M_j = a_j X + w_j W + g_j WX$$
$$Y = b_j M_j.$$

(1)

Substitute M_j:

$$Y = b_j(a_j X + w_j W + g_j WX)$$
$$Y = b_j w_j W + (a_j b_j + b_j g_j W)X.$$

(2)

A quantification the effect of X on Y as a function of the moderator has been described by Hayes (2015, 2018), who coined the term "index of moderated mediation". The index of moderated mediation (IMM) is defined here as:

$$IMM_a = a_j b_j + b_j g_j W.$$

(3)

This definition is slightly different from that given by Hayes (2105), since contrary to Hayes(2015), here it also contains the term constant term $a_j b_j$. The IMM for a single path represents the slope (and intercept) of the line that relates the predictor with the dependent variable, as a function of the moderator. The index consists of a fixed part (the indirect effect) and a part that varies with the moderator. The plot of the IMM versus the

moderator can be constructed using formula (2), which gives a straight line. This plot shows how the relation between predictor and dependent variable changes when the moderator changes. A flat horizontal line in this plot indicates that there is no moderation. A steep line om the other hand indicates a moderation effect. When $W = 0$ the IMM represents the indirect (unmoderated) effect. The single-path IMM is defined for each mediator separately. Likewise, the IMM is defined for each path separately, thus for the path from predictor to mediator (a-path) and also for mediator to the dependent variable (b-path). Foth this reason we refer to it as a single-path IMM. So, for three mediators and two moderators (one for each path), six single-path IMMs can be computed. Combining both moderators to obtain a double-path IMM would give a two-dimensional plane in a three-dimensional space (instead of a line), defined by the two moderators and the IMM. The index is derived as follows.

$$M_j = a_j X + w_j W + g_j W X$$
$$Y = b_j M_j + vV + h_j M_j V. \tag{4}$$

Substitute M_j:

$$Y = b_j(a_j X + w_j W + g_j W X) + vV + h_j(a_j X + w_j W + g_j W X)V$$
$$Y = b_j w_j W + (a_j b_j + b_j g_j W)X + vV + h_j w_j WV + (h_j a_j V + h_j g_j WV)X \tag{5}$$
$$Y = b_j w_j W + vV + g_j w_j WV + (a_j b_j + b_j g_j W + h_j a_j V + h_j g_j WV)X.$$

The two-dimensional IMM, incorporating both moderators is:

$$IMM = a_j b_j + b_j g_j W + h_j a_j V + h_j g_j WV. \tag{6}$$

When one of the moderators is dichotomous, it is more convenient to plot the double-path IMM in a two dimensional space with two lines, one for each category of the

moderator. It is easy to see that formula (6) is a generalization of formula (3), when V = 0 (or absent), and when W = 0 (or absent), the index of moderated mediation (IMM) of the b-path is defined as:

$$IMM_b = a_j b_j + h_j a_j V. \tag{7}$$

Like before, this index indicates how the slope of the effect of X on Y changes as a function of the moderator, here the moderator of the b-path (V)

Examples

The required parameters of the function `gemm()` can be found by typing `?gemm`. This gives the help page of the function. The fictious data that will be used to illustrate the model are included in the package and are in the R object `gemmDat`, which can be loaded by `data("gemmDat")`. After loading the data, you can also learn more about the data by using the question mark.

```
data("gemmDat")            # loads the data
```

The data have simple and commonly used variable names: x and y for the predictor and dependent variable, respectively. The "m" with a number represents mediators and the "c" covariates, numerical moderators start with "mod", and dichotomous moderators with "bi".

The mediation model

The first model that will be illustrated is a straightforward mediation model with three mediators, without moderators and without covariates. The number of bootstrap samples is set to 100, a very small number for bootstraps, but here only used for

illustration. For a real analysis you might use 5,000 boostraps; the default for nboot is

1,000.

```
result <- gemm(dat = gemmDat,

               xvar ="x1",

               mvars = c("m1","m2","m3"),

               yvar = "y1",

               nboot = 50)
```

The parameter "xvar" represents the predictor, "yvar" the dependent variable and

"mvars" a vector of mediators. The result of the analysis is put in the object "result".

Except for nboot, these parameters are all obligatory to specify. This implies that neither a

simple regression model, nor a moderation model, can be analyzed with this function.

After running the analysis, the results can be viewed by typing:

```
print(result)
```

```
## ###    The model contains 3 mediators: m1 m2 m3

##         and 0 moderators for the x-m path(s):

##         and 0 moderators for the m-y path(s):

##         and 0 covariates for the mediators:

##         and 0 covariates for the dependent variable:

##

## ### Explained variance (R-square) of the mediators and dependent variable:

##

##     m1    m2    m3    y1

## 0.917 0.053 0.243 0.826

##

## ### Direct effect:
```

```
##
##    label    est    se     z pvalue ci.lower ci.upper
## 7       c 0.006 0.02 0.32    0.75    -0.035     0.048
##
## ### Indirect effects (unstandardized):
##
##     label      est     se        z pvalue ci.lower ci.upper
## 16   ind1   0.130 0.019   6.959   0.000     0.085     0.173
## 17   ind2  -0.056 0.023  -2.427   0.015    -0.110    -0.019
## 18   ind3   0.000 0.004  -0.035   0.972    -0.008     0.010
##
## ### Indirect effects (standardized):
##
##            ind
## m1   0.47429
## m2  -0.20404
## m3  -0.00049
```

The printed output consists of five parts. First, the function input is shown. Here you can check whether you have specified the model that you intended to run. Second, the explained variance is shown for the dependent variable and the mediators. For the dependent variable the variance is explained by the predictor, the mediator(s), and optionaly also by the covariate(s).For the mediators the variance is explained by the predictor, and optionaly also by the covariate(s). The third part shows the direct effect of the predictor on the dependent variable. The confidence interval around this estimate has been obtaind by using the bootstrap samples. The fourth part shows the indirect effects. Each row represents an indirect effect through a particular mediator. The last row is the total of all indirect effects. Again boostrap samples are used to obtain the estimates.

Finally, in part five the standardized indirect effects are given, one for each mediator. They represent the completely standardized effect size of the indirect effect. This is the effect size, obtained after all variables have been standardized.

The moderated mediation model

In the next example a moderated mediation model is shown.

```r
result <- gemm(dat = gemmDat,

               xvar  = "x1",

               mvars = c("m1","m2"),

               yvar  = "y1",

               mymod = "bimod2",

               cmvars = c("c1","c2"),

               cyvars = c("c1","c2"),

               nboot = 50)


print(result)
```

```
## ###    The model contains 2 mediators: m1 m2

##         and 0 moderators for the x-m path(s):

##         and 1 moderators for the m-y path(s): bimod2

##         and 2 covariates for the mediators: c1 c2

##         and 2 covariates for the dependent variable: c1 c2

##

## ### Explained variance (R-square) of the mediators and dependent variable:

##

##     m1     m2     y1

## 0.874 0.076 0.838
```

```
##
## ### Direct effect:
##
##   label    est    se    z pvalue ci.lower ci.upper
## 5      c 0.005 0.022 0.21   0.83   -0.047    0.051
##
## ### Indirect effects (unstandardized):
##
##    label    est    se    z pvalue ci.lower ci.upper
## 36  ind1  0.13 0.024  5.4   0.00    0.086     0.19
## 37  ind2 -0.03 0.024 -1.2   0.21   -0.089     0.03
##
## ### Indirect effects (standardized):
##
##      ind
## m1  0.47
## m2 -0.11
```

Here we have moderator ("bimod2") added to the model. This variable is assumed to
moderate the two effects of the two mediators on the dependent variable. This moderator is
a dichotomous moderator, in R terminology a factor with two levels (e.g. male and female).
The variables c1 and c2 are used as covariates. They are used two times, first as covariates
of the mediators (cmvars) and also as covariates of the dependent variabele (cyvars). The
structure of the printed output is similar to the previous analyses. However, we can ask for
plots now, because the model contains a moderator. There actually are three plot types.
The first type is the single-path index of moderated mediation (Hayes, 2018, p.425),
plotted against the moderator. As explained before, the index of moderated mediation
(IMM) is the slope of the line representing the effect of the mediator (c.q. predictor) on the

dependent variable (c.q. mediator) that changes when the moderator changes. When there are two mediators in the model there are also two IMMs. A flat horizonal line indicates that the slope does not change as a function of the supposed moderator, in that case there is no moderation. The steeper the line the stronger the moderated mediation effect. For a dichotomous moderator only two points are shown (per mediator) with their 95% confidence interval. The command to obtain the single-path IMM is `plotIMM(result)`.

plotIMM(result)
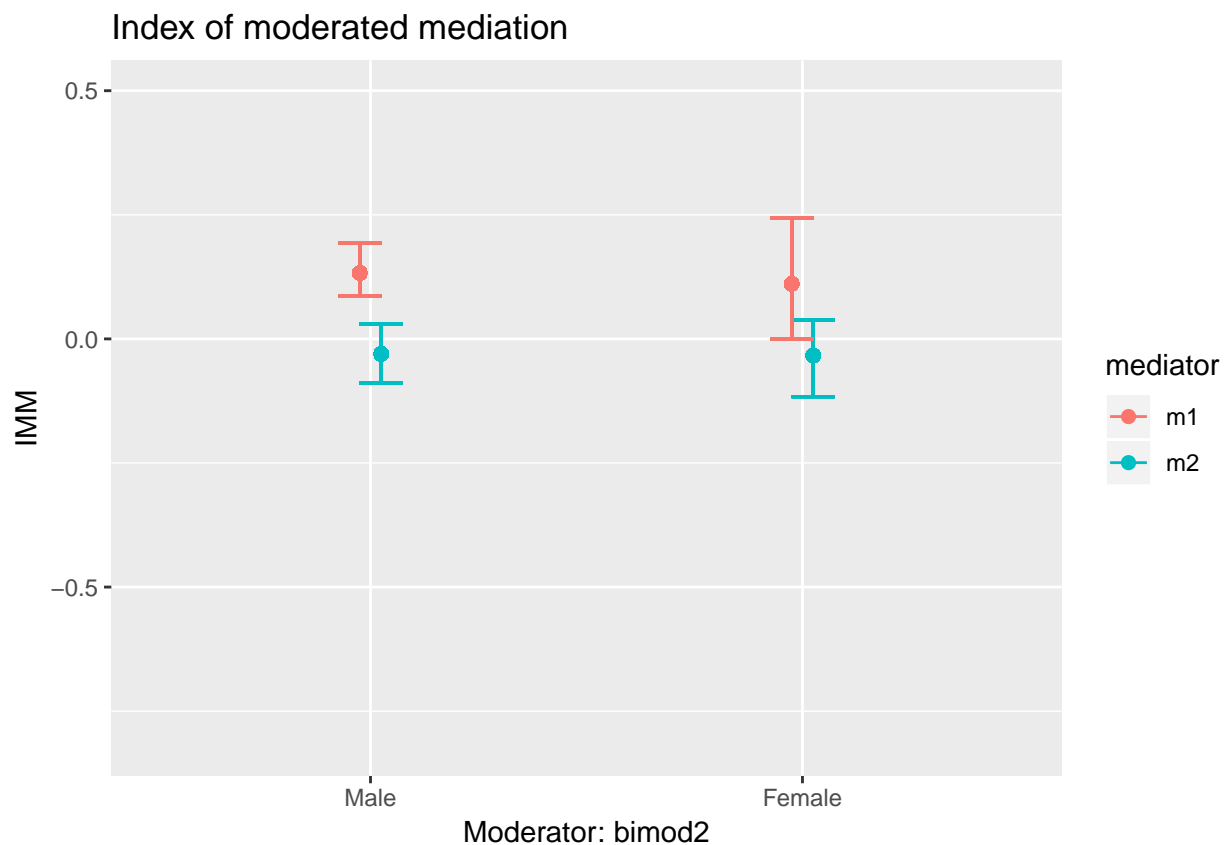
The IMM for the first mediator (m1) is somewhat larger than for the second (m2). However, for both mediators the IMM hardly changes between the two categories of the moderator.

The second plot type is the mediated simple slopes plot. A simple slope plot shows the indirect effect of the predictor on the dependent variable through a mediator for

usually two characteristic values of the moderator. Each value of the moderator is represented by a separate line. Parallel lines indicate absence of moderation and crossing lines the presence of moderation. The 95% CI interval around each line is indicated by a shaded area. For each mediator a separate subplot is given.
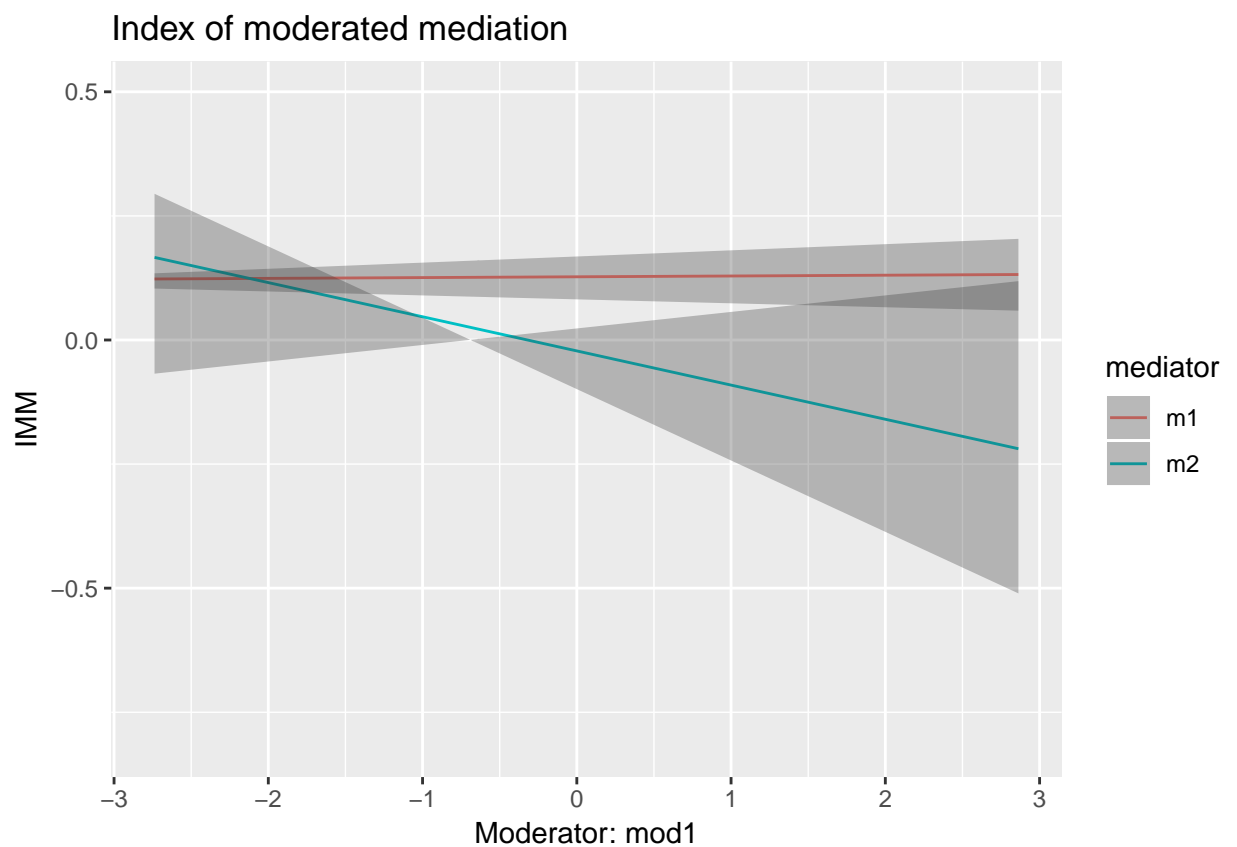
```
plotSS(result)
```


Simple slopes in m–y path for indirect effect
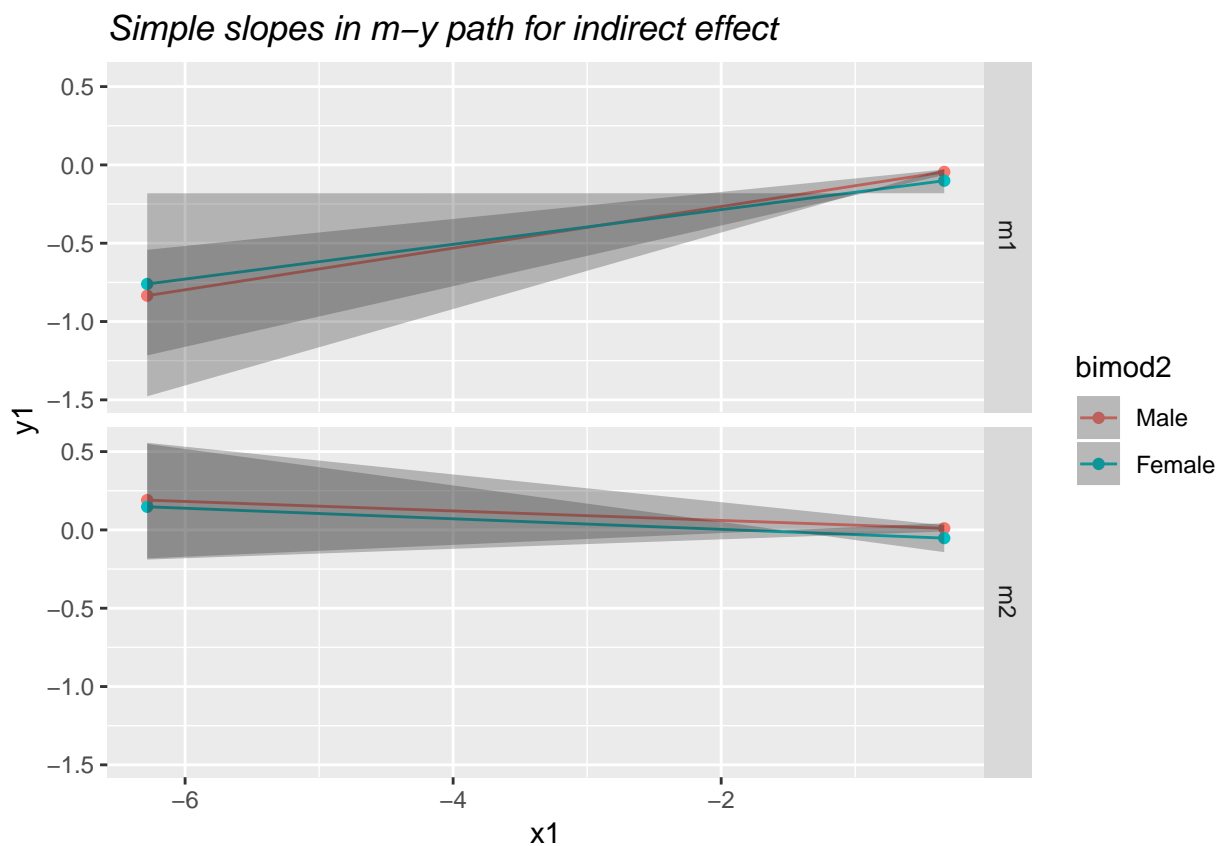
These plots show the effect of the predictor on the dependent variable through the mediators m1 and m2, respectively. Because the moderator is a dichotomous variable each plot contains two lines, one for each category of the moderator.
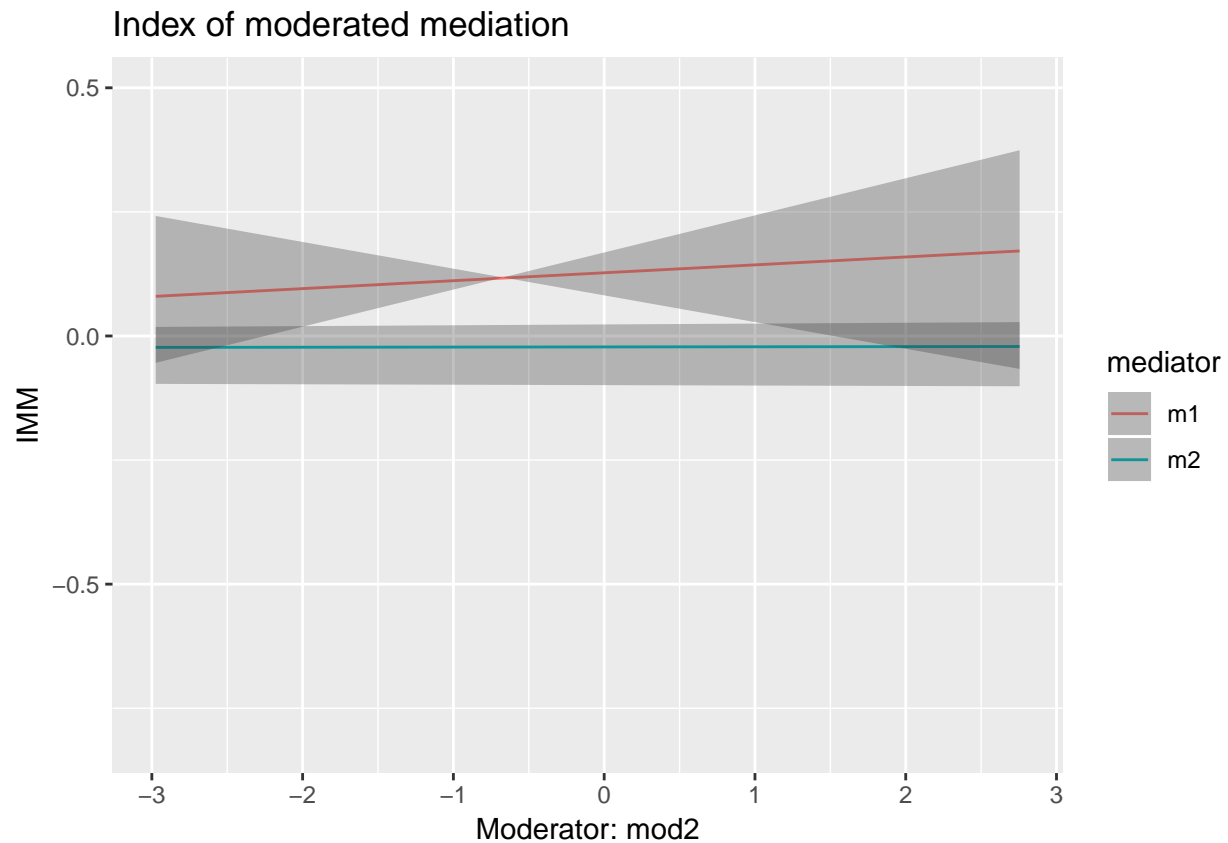
As a final example we illustrate the use of two numerical moderators.

```
result <- gemm(dat = gemmDat,
               xvar  = "x1",
```

```
          mvars = c("m1","m2"),

          yvar  = "y1",

          xmmod = "mod1",

          mymod = "mod2",

          nboot = 50)


plotIMM(result)
```

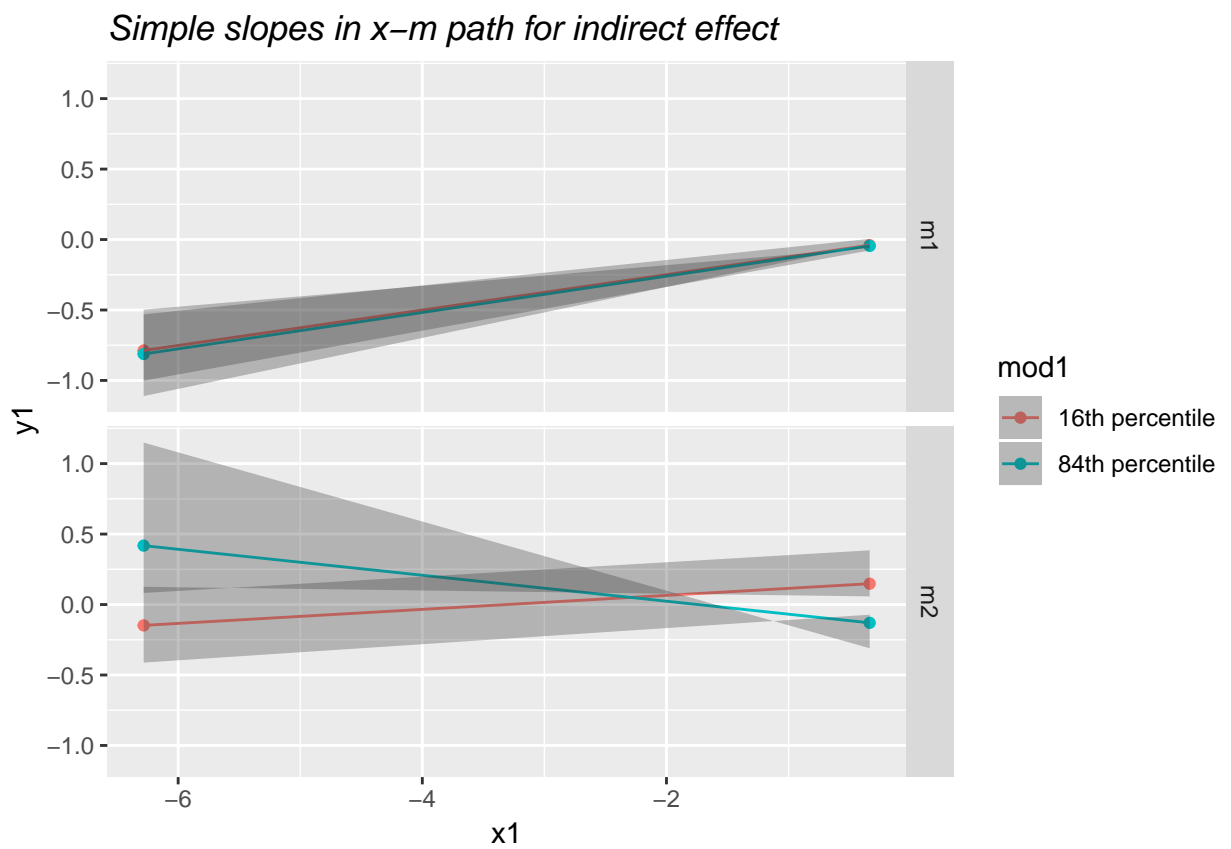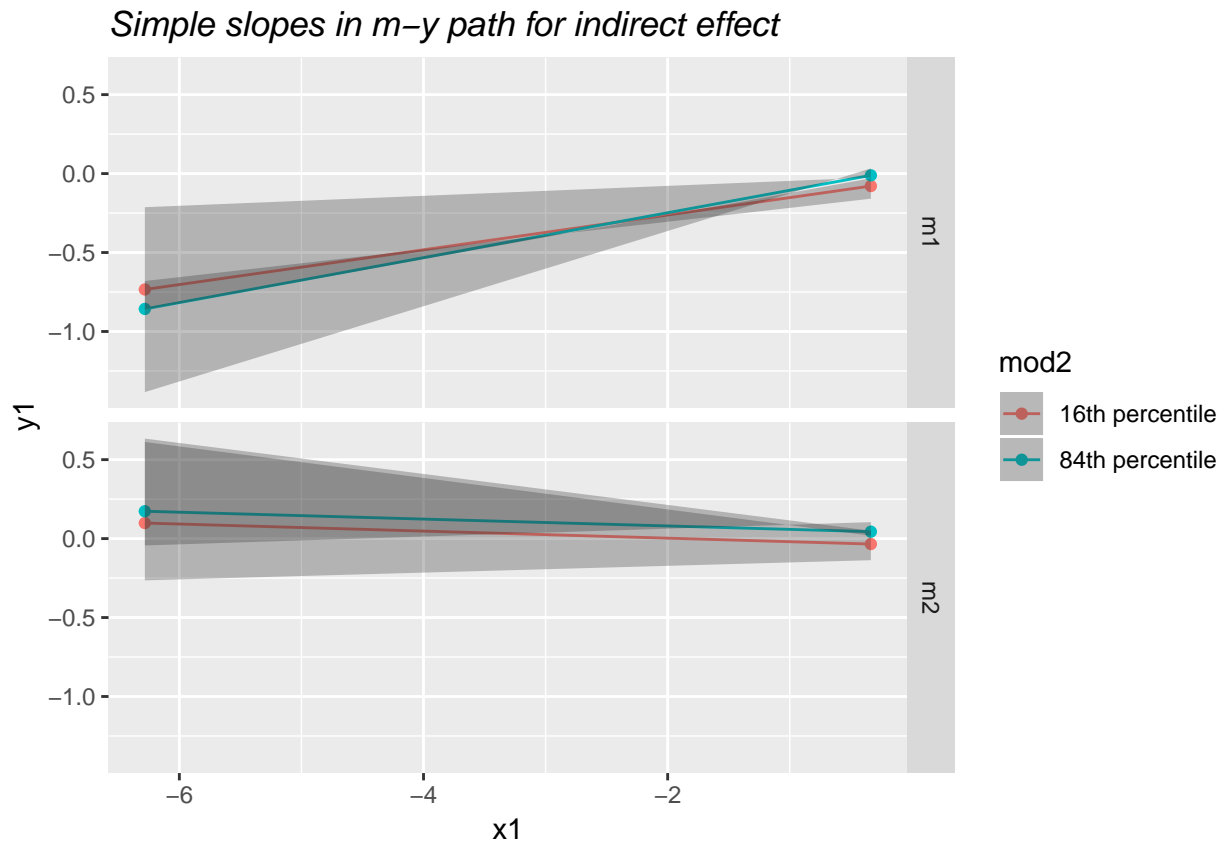### Index of moderated mediation

Index of moderated mediation

The first plot command provides two plots, for each moderator a separate plot of the single-path IMM, with in each plot a separate line for each mediator. In the first plot (for mod1) we see that for the second mediator the IMM becomes more negative as the value of the moderator increases. So, the effect of x1 on y1 is more negative for larger values of mod1.

The second plot command provides a simple slopes plots, one for each moderator, where each plot is split into subplots for each mediator. The lines represent the effect evaluated at the 16th and 84th percentile of the moderator, as recommended by Hayes(2018). Only for mod1 and mediator m2 there seems moderation, because we can see crossing lines. In the other three cases the lines run parallel to each other, indicating no relevant moderation effects.

```
plotSS(result)
```

Simple slopes in x–m path for indirect effect

Simple slopes in m–y path for indirect effect

The `plotIMM3d(result)` command provides 3d-plots showing the double-path IMM for both moderators. Separate plots are given for each mediator. The message is the same as in the previous plots: only the moderated mediation effect appears to be present: the indirect effect of x1 on y through the second mediator is moderated by mod1.

```
plotIMM3d(result)
```

Index of Moderated Mediation of ab−path for mediator: m1
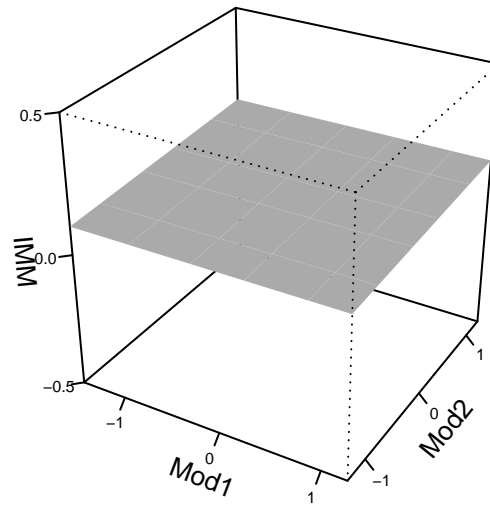
Index of Moderated Mediation of ab–path for mediator: m2
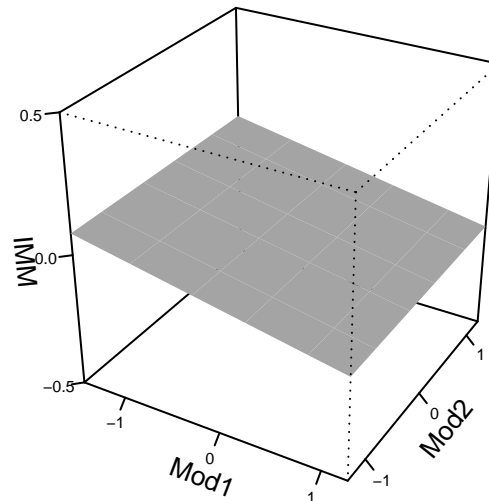
Additional output

The `result` object is a list of three elements, which are named input, intermediate, and results, respectively. Input contains the original data and all variable names used in the analysis. Intermediate contains results that have been computed in the functions. For instance the object `result$intermediate$model` contains the model specification that was used by lavaan, the object `result$intermediate$result` contains the output of lavaan. You can inspect this output by loading the package lavaan and then using `summary(result$intermediate$result)`. All other lavaan extract functions can be used on this object. The object `result$intermediate$parameterEstimates` contains all estimated parameters from lavaan. For example, when you want to see some important fit measures obtained in lavaan, you may type:

```
lavaan::fitmeasures(result$intermediate$result)[c("cfi","tli","rmsea","chisq", "df")]
```

```
##      cfi      tli    rmsea    chisq       df
##    0.502   -0.087    0.630 1321.917   11.000
```

In this way you are able to inspect all other lavaan results.

Note

We used R [Version 3.5.1; @] and the R-packages * base* [@ R-base], *gemm* (Version 0.0.1; Verboon, 2018), and *papaja* (Version 0.1.0.9842; Aust & Barth, 2018) for all our analyses.

References

Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown.* Retrieved from https://github.com/crsh/papaja

Edwards, J. R., & Lambert, L. S. (2007). Methods for integrating moderation and mediation: A general analytical framework using moderated path analysis. *Psychological Methods*, *12*, 1–22.

Hayes, A. F. (2015). An Index and Test of Linear Moderated Mediation. *Multivariate Behavioral Research*, *50*(1), 1–22. https://doi.org/10.1080/00273171.2014.962683

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis.* NY: Guilford Press.

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Assessing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research*, *42*, 185–227.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. Retrieved from http://www.jstatsoft.org/v48/i02/

Verboon, P. (2018). *Gemm: Moderated mediation with covariates.*