

1 Moderation of the regression of m on x

Consider first the m, x, z part of the model. The mediator regression

$$m_i = \gamma_0 + \gamma_{1i} x_i + \epsilon_{1i}, \quad (1)$$

allows a moderated effect of x on m using a slope γ_{1i} that varies with a moderator variable z_i ,

$$\gamma_{1i} = \gamma_1 + \gamma_2 z_i, \quad (2)$$

Inserting γ_i in the mediator regression of (1) results in

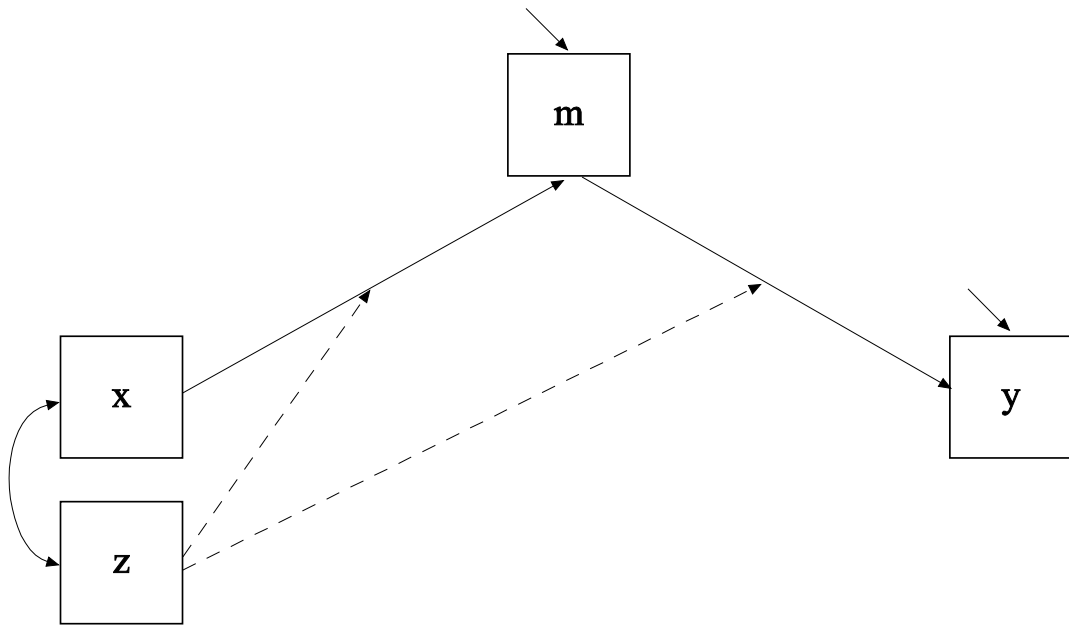
$$m_i = \gamma_0 + (\gamma_1 + \gamma_2 z_i) x_i + \epsilon_{1i}, \quad (3)$$

$$= \gamma_0 + \gamma_1 x_i + \gamma_2 z_i x_i + \epsilon_{1i}. \quad (4)$$

In (3) the moderation effect is expressed as the moderator function $(\gamma_1 + \gamma_2 z_i)$. The main effect of x is γ_1 and the interaction effect is γ_2 . The moderator variable z can be continuous or binary. In the binary case, an alternative way to include moderation is to carry out a multiple-group analysis.

The γ_{1i} slope in (2) is a dependent variable in the regression on z. The γ_{1i} regression can also include a residual, giving rise to the term random slope, emphasizing that z may not be the only determinant of γ_{1i} . In (2) the residual is zero, which is equivalent to adding an interaction term $z \times x$ as a covariate in the regression of m as seen in (4). The simulations to be reported use the approach

Figure 1: Moderation and mediation



of (1) and (2) to create this interaction.

When ignored, a non-zero γ_{1i} residual leads to a heteroscedastic residual in the regression of m on x and $z \times x$, that is, a residual that varies with the x values. This produces misestimated standard errors. Mplus can analyze both the case of a zero and non-zero residual in line with random coefficient estimation in econometrics (see User's Guide example 3.3 and the references given there). It is unknown how common the need is for a non-zero residual. Simulations suggest that when it is present and ignored, the distortion of standard errors may not be critical.

Consider the model of Figure 2 using a filled circle to denote the interaction between z and x . For simplicity, no effect of z on m is included although this would typically be the case in practice. An example of this type of model is

given in Figure ?? discussed in Section ?. Table 1 shows the input for a Monte Carlo simulation with $n = 400$ and using 500 replications. This means that 500 samples of size 400 are generated. Data are generated according to the model given in the MODEL POPULATION command. The variables are normally distributed, except that x is dichotomized at its mean using the CUTPOINTS option. The covariates x and z are uncorrelated. In this way, the data can be thought of as coming from a randomized treatment study, where x is the 0/1 treatment variable. The MODEL command specifies that all the 500 data sets should be analyzed by the same model as given in MODEL POPULATION. Note that the covariate part of MODEL POPULATION is not specified in the MODEL command because parameters for covariates are not estimated. The MODEL statements use parameter labels in parentheses, corresponding to the notation in (1) and (2). The random slope γ_{1i} of (1) is specified using the bar (|) notation. The bracket expression labeled (gamma1) is the γ_1 intercept in the γ_{1i} expression of (2). The MODEL CONSTRAINT command expresses the moderated indirect effect of x on y via m , $\gamma_1 + \gamma_2 z_i$, at the average value of z and at minus and plus one standard deviation from the mean of z . The mean and variance of z are given as zero and one in the MODEL POPULATION command.

Figure 2: Moderation of the regression of m on x using interaction notation

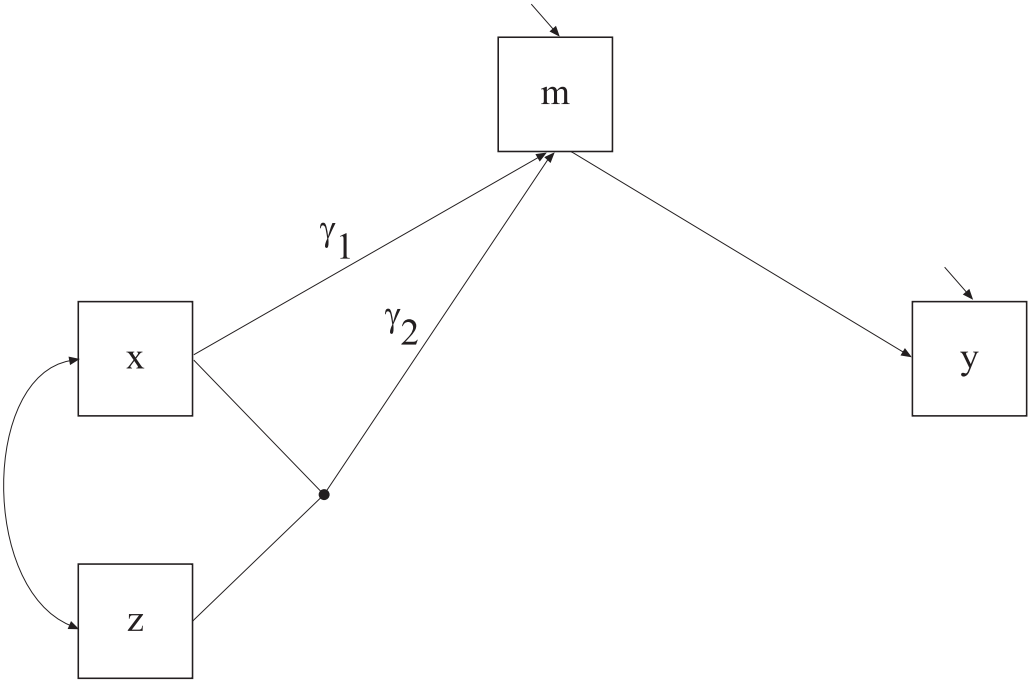


Table 1: Input for Monte Carlo simulation of z moderation of x to m

```

TITLE:      Simulating Z moderation of X to M using a random slope,
             saving the data for external Monte Carlo analysis
MONTECARLO:
            NAMES = y m x z;
            NOBS = 400;
            NREPS = 500;
            REPSAVE = ALL;
            SAVE = zxrep*.dat;
            CUTPOINTS = x(0);
MODEL POPULATION:
            x-z@1;
            x WITH z@0;
            y ON m*.5;
            gamma1 | m ON x;
            [gamma1*.3];
            gamma1 ON z*.2;
            gamma1@0;
            y*.5;
            m*1;
ANALYSIS:  TYPE = RANDOM;
MODEL:     y ON m*.5 (b);
            gamma1 | m ON x;
            [gamma1*.3] (gamma1);
            gamma1 ON z*.2 (gamma2);
            gamma1@0;
            y*.5;
            m*1;
MODEL CONSTRAINT:
            NEW(indavg*.15 indlow*.05 indhigh*.25);
            indavg = b*gamma1;
            indlow = b*(gamma1-gamma2);
            indhigh = b*(gamma1+gamma2);

```
