Mplus Workshop, Part 1: Highlights from Muthén, Muthén & Asparouhov (2016) Regression And Mediation Analysis Using Mplus

Bengt Muthén
bmuthen@statmodel.com

Mplus
www.statmodel.com

Pre-conference workshop
at the Modern Modeling Methods Conference
UConn, May 23, 2016

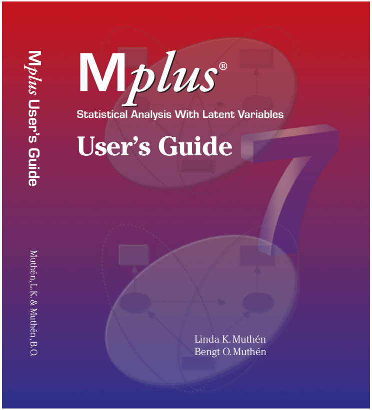
Expert assistance from Noah Hastings is acknowledged

Mplus Workshop: Overview of the Day

- 8:30 - 12:00: Highlights from the new book (Mplus Version 7.4)
  - First morning block (Bengt, 1 1/2 hours[1]): Regression analysis
  - Second morning block (Bengt, 1 1/2 hours): Mediation analysis
- Lunch: 12 - 1:30
- 1:30 - 6:30 (or longer): Time-series analysis (forthcoming Mplus Version 8)
  - First afternoon block (Ellen, 1 1/2 hours): Introductory time-series analysis
  - Second afternoon block (Ellen, 1 1/2 hours): Examples
  - Third afternoon block (Tihomir, 1 1/2 hours): Time-series implementation in Mplus Version 8

[1] 10-15 minutes of questions and answers at the end of each block (hold your questions).

The Mplus User's Guide has Gotten a Companion



Chapters of Regression And Mediation Analysis Using Mplus

- 1. Linear regression analysis
- 2. Mediation analysis
- 3. Special topics in mediation analysis
- 4. Causal inference for mediation
- 5. Categorical dependent variable
- 6. Count dependent variable
- 7. Censored variable
- 8. Mediation with non-cont's variables
- 9. Bayesian analysis
- 10. Missing data

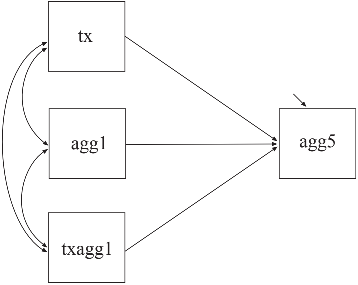
Table of Contents will be shown at www.statmodel.com. 500 pages. Lots of inputs and outputs. Paperback. All inputs and outputs will be posted. Most data sets will be posted. Perhaps assignments.

Overview of the Morning: Highlights from the Book

- First morning block (1 1/2 hours). Regression Analysis:
  - Linear regression with an interaction
  - Heteroscedasticity modeling
  - Censored variable modeling: Tobit, censored-inflated, Heckman, and two-part analysis
  - Bayes: Advantages over ML. Missing data on covariates
- Second morning block (1 1/2 hours). Mediation Analysis:
  - Moderated mediation with continuous mediator and outcome
  - Monte Carlo simulation of moderated mediation
  - Sensitivity analysis
  - Mediation analysis using counterfactually-defined indirect and direct causal effects:
    - Binary outcome
    - Count outcome
    - Two-part outcome

Note: The highlights skew toward the more advanced parts of the book to match the claim ”Analyses you probably didnt know that you could do in Mplus”.

Example: Linear Regression with an Interaction



Randomized field experiment in the Baltimore public schools where a classroom-based intervention aimed at reducing aggressive-disruptive behavior among elementary school students was carried out (Kellam et al., 2008)

- tx is a binary intervention variable
- agg1 is pre-intervention Grade 1 aggressive behavior score and agg5 the score in Grade 5
- txagg1 is a treatment-baseline interaction (tx $\times$ agg1)

Example: Linear Regression with an Interaction

$$agg5_i = \beta_0 + \beta_1\ tx_i + \beta_2\ agg1_i + \beta_3\ txagg1_i + \varepsilon_i. \tag{1}$$

$$\begin{aligned} agg5_i &= \beta_0 + \beta_1\ tx_i + \beta_2\ agg1_i + \beta_3\ tx_i\ agg1_i + \varepsilon_i & (2)\\ &= \beta_0 + \beta_2\ agg1_i + (\beta_1 + \beta_3\ agg1_i)\ tx_i + \varepsilon_i. & (3)\end{aligned}$$

The expression $\beta_1 + \beta_3\ agg1$ is referred to as the moderator function or, when evaluated at a specific agg1 value, the simple slope. This means that agg1 moderates the β_1 effect of tx on agg5 by the term $\beta_3\ agg1$.

Example: Input for Linear Regression with an Interaction

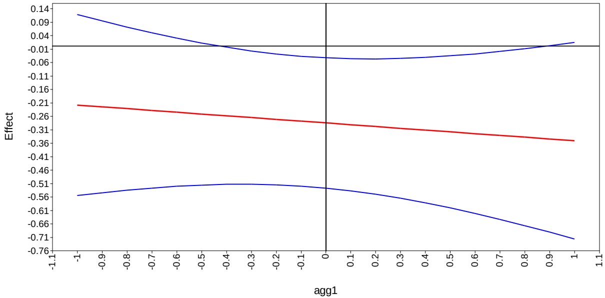
```
VARIABLE:          USEVARIABLES = agg5 agg1 tx txagg1;
                   USEOBSERVATIONS = gender EQ 1 AND (desgn11s EQ 1 OR
                   desgn11s EQ 2 OR desgn11s EQ 3 OR desgn11s EQ 4);
DEFINE:            IF (desgn11s EQ 4) THEN tx=1;
                   IF (desgn11s EQ 1 OR desgn11s EQ 2 OR desgn11s EQ 3) THEN
                   tx=0;
                   agg5 = sctaa15s;
                   agg1 = sctaa11f;
                   CENTER agg1(GRANDMEAN);
                   txagg1 = tx*agg1;
ANALYSIS:          ESTIMATOR = MLR;
MODEL:             agg5 ON
                   tx (b1)
                   agg1 (b2)
                   txagg1 (b3);
MODEL CONSTRAINT:
                   NEW(modlo mod0 modhi);
                   modlo = b1+b3*(-1.06);
                   mod0 = b1;
                   modhi = b1+b3*1.06;
OUTPUT:            SAMPSTAT PATTERNS STANDARDIZED RESIDUAL TECH4;
PLOT:              TYPE = PLOT3;
```

Table : Results for regression with a randomized intervention using treatment-baseline interaction ($n = 250$)

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
agg5 ON				
tx	-0.285	0.124	-2.307	0.021
agg1	0.500	0.076	6.543	0.000
txagg1	-0.066	0.130	-0.511	0.609
Intercepts				
agg5	2.483	0.077	32.238	0.000
Residual variances				
agg5	0.952	0.090	10.612	0.000
New/additional parameters				
modlo	-0.215	0.177	-1.211	0.226
mod0	-0.285	0.124	-2.307	0.021
modhi	-0.355	0.192	-1.849	0.064

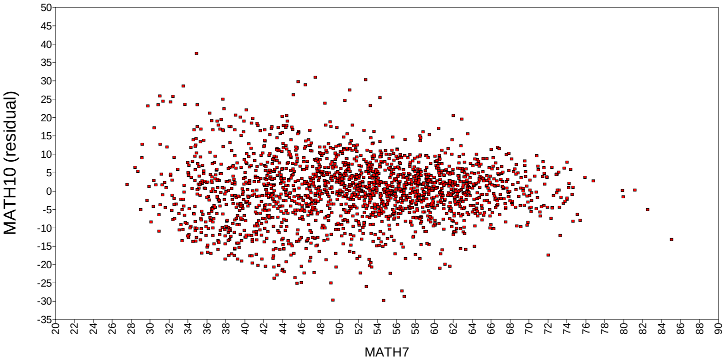
Example: Linear Regression with an Interaction (Alt.)

```
MODEL:    agg5 ON
          tx (b1)
          agg1 (b2)
          txagg1 (b3);
MODEL CONSTRAINT:
          LOOP(x,-1,1,0.1);
          PLOT(effect);
          effect = b1+b3*x;
```



Heteroscedasticity Modeling: Example: LSAY Math Data ($n = 2,019$)

Figure : Linear regression residuals for math10 plotted against math7



Heteroscedasticity Modeling: (1) Using MODEL CONSTRAINT

The linear regression model assumes homoscedastic residual variances,

$$y_i = \beta_0 + \beta_1\, x_i + \varepsilon_i, \tag{4}$$

$$V(\varepsilon_i | x_i) = V(\varepsilon_i) = V(\varepsilon). \tag{5}$$

An exponential function may instead be used for the residual variance,

$$V(\varepsilon_i | x_i) = e^{a + b\, x_i}, \tag{6}$$

where a and b are parameters to be estimated. If $b = 0$, $V(\varepsilon_i | x_i) = e^a$ which means that the residual variance is not a function of x so that homoscedasticity holds. If $b > 0$, the residual variance increases as a function of x and if $b < 0$, the residual variance decreases as a function of x.

Input for Heteroscedasticity Modeling

TITLE:	Regressing math10 on math7 with heteroscedasticity
DATA:	FILE = dropout.dat;
	FORMAT = 11f8 6f8.2 1f8 2f8.2 10f2;
VARIABLE:	NAMES = id school gender mothed fathed fathsei ethnic expect pacpush pmpush homeres math7 math8 math9 math10 math11 math12 problem esteem mathatt clocatn dlocatn elocatn flocatn glocatn hlocatn ilocatn jlocatn klocatn llocatn;
	MISSING = mothed (8) fathed (8) fathsei (996 998) ethnic (8) homeres (98) math7-math12 (996 998);
	IDVARIABLE = id;
	USEVARIABLES = math7 math10 mothed male;
	CONSTRAINT = math7;
DEFINE:	male = gender - 1;
ANALYSIS:	**STARTS = 10;**
	BOOTSTRAP = 1000;
MODEL:	math10 ON math7 mothed male;
	math10 (resvar);
MODEL CONSTRAINT:	
	NEW(a b);
	resvar = EXP(a+b*math7);
OUTPUT:	TECH8 SAMPSTAT
	CINTERVAL(BOOTSTRAP);
PLOT:	TYPE = PLOT3;

Table : Loglikelihood and BIC for heteroscedasticity modeling of LSAY math data

	#par's	logL	BIC
Regular regression	5	-6972	13982
Heteroscedasticity regression	6	-6885	13816

Non-Symmetric Bootstrap Confidence Intervals for Heteroscedasticity Modeling of the LSAY Math Data

	Lower 2.5%	Lower 5%	Estimate	Upper 5%	Upper 2.5%
math10 ON					
math7	0.981	0.986	1.017	1.050	1.058
mothed	0.529	0.579	0.872	1.148	1.192
male	0.115	0.215	0.822	1.426	1.514
Intercepts					
math10	6.664	7.061	8.759	10.444	10.804
Residual Variances					
math10	999.000	999.000	999.000	999.000	999.000
New/Additional Parameters					
a	6.020	6.086	6.379	6.704	6.761
b	-0.051	-0.050	-0.043	-0.037	-0.036

- Assuming homoscedasticity: Non-significant effect of male, 95% CI is $[-0.167, 1.336]$
- Allowing heteroscedasticity: Significant effect of male, 95% CI is $[0.115, 1.514]$

Heteroscedasticity Modeling: (2) Using Random Coefficients

$$y_i = \beta_0 + \beta_{1i}\, x_i + \beta_2\, z_i + \varepsilon_i, \tag{7}$$

$$\beta_{1i} = \beta_1 + \beta_3\, z_i + \delta_i. \tag{8}$$

The residuals ε and δ are allowed to covary. The model can be compared to regular regression with an interaction between the covariates x and z by inserting (8) into (7),

$$y_i = \beta_0 + \beta_1\, x_i + \beta_3\, x_i\, z_i + \delta_i\, x_i + \beta_2\, z_i + \varepsilon_i. \tag{9}$$

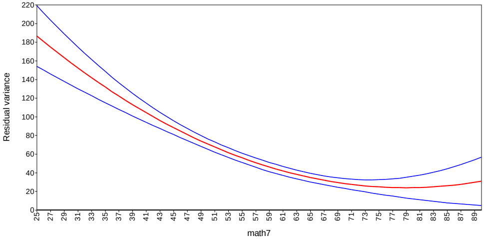
The random coefficient model allows for a heteroscedastic residual variance. Whereas in regular regression the residual variance is assumed to be the same for all individuals, $V(y \mid x, z) = V(\varepsilon)$, the residual variance for the random coefficient model varies with x. The conditional variance of y in (9) is

$$V(y_i \mid x_i, z_i) = V(\delta_i)\, x_i^2 + 2\, Cov(\delta_i, \varepsilon_i)\, x_i + V(\varepsilon_i). \tag{10}$$

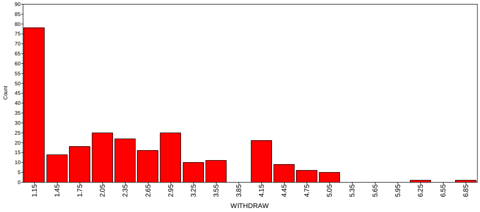
Heteroscedasticity Using Random Coefficients

```
ANALYSIS:          TYPE = RANDOM;
MODEL:             s |math10 ON math7;
                   s WITH math10 (cov);
                   math10 (resvary);
                   s (vbeta);
OUTPUT:            TECH1 SAMPSTAT STDYX RESIDUAL CINTERVAL;
PLOT:              TYPE = PLOT3;
MODEL CONSTRAINT:
                   PLOT (vygivenx);
                   LOOP(x,25,90,1);
                   vygivenx = vbeta*x*x + 2*cov*x + resvary;
```

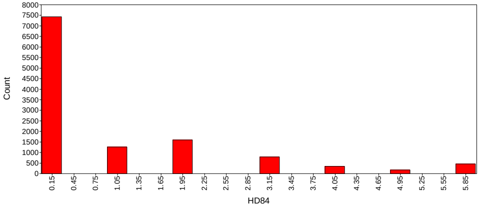
- Better BIC than homoscedastic model



30% floor effect:



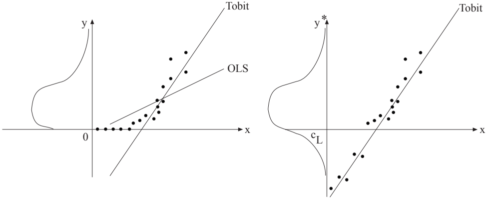
59% floor effect:



Regression Analysis Options in Mplus

- Censored-normal (Tobit)
- Censored-inflated
- Sample selection (Heckman)
- Two-part

Censored-Normal (Tobit) Regression



$$y_i^* = \beta_0 + \beta_1 \, x_i + \varepsilon_i, \tag{11}$$

$$y_i = \begin{cases} 0 & \text{if } y_i^* \leq 0 \\ y^* & \text{if } y_i^* > 0 \end{cases}$$

$$Binary\ (probit): \ P(y_i > 0 | x_i) = 1 - \Phi[\frac{0 - \beta_0 - \beta_1 x_i}{\sqrt{V(\varepsilon)}}] = \Phi[\frac{\beta_0 + \beta_1 x_i}{\sqrt{V(\varepsilon)}}], \tag{12}$$

$$Continuous, positive: \ E(y_i | y_i > 0, x_i) = \beta_0 + \beta_1 \, x_i + \sqrt{V(\varepsilon)} \, \frac{\phi(z_i)}{\Phi(z_i)}, \tag{13}$$

- Latent class 0: subjects for whom only $y = 0$ is observed
- Latent class 1: subjects following a censored-normal (tobit) model

Assume a logistic regression that describes the probability of being in class 0,

$$logit(\pi_i) = \gamma_0 + \gamma_1\, x_i. \tag{14}$$

For subjects in class 1 the usual censored-normal model of (15) is assumed with

$$y_i^* = \beta_0 + \beta_1\, x_i + \varepsilon_i. \tag{15}$$

Two ways $y = 0$ is observed (mixture at zero).

Sample Selection (Heckman) Regression

Consider the linear regression for the continuous latent response variable y^*,

$$y_i^* = \beta_0 + \beta_1\, x_i + \varepsilon_i, \tag{16}$$

where the latent response variable y_i^* is observed as $y_i = y_i^*$ when a binary variable $u_i = 1$ and remains latent, that is, missing if $u_i = 0$. A probit regression is specified for u,

$$u_i^* = \gamma_1\, x_i + \delta_i, \tag{17}$$

where the categories of the binary observed variable u_i are determined by u^* falling below or exceeding a threshold parameter τ,

$$u_i = \begin{cases} 0 & \text{if } u_i^* \leq \tau \\ 1 & \text{if } u_i^* > \tau. \end{cases}$$

A key feature is that the residuals ε and δ are assumed to be correlated and have a bivariate normal distribution with the usual probit standardization $V(\delta) = 1$.

Two-Part Regression

With censoring from below at zero and using probit regression with the event of $u = 1$ referring to a positive outcome, the two-part model is expressed as

$$probit(\pi_i) = \gamma_0 + \gamma_1\, x_i, \tag{18}$$

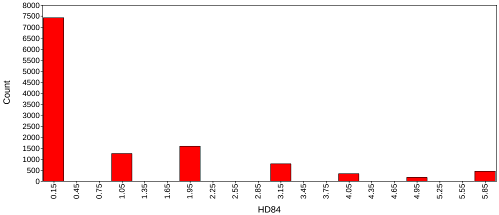
$$log\ y_{i|u_i=1} = \beta_0 + \beta_1\, x_i + \varepsilon_i, \tag{19}$$

where $\pi_i = P(u_i = 1|x_i)$ and $\varepsilon_i \sim N(0, V(\varepsilon))$. Logistic regression can be used as an alternative to the probit regression in (18). Maximum-likelihood estimation of the two-part model gives the same estimates as if the binary and the continuous parts were estimated separately using maximum-likelihood. Expressing (18) in terms of a latent response variable regression with a normal residual, the two residuals can be correlated but the correlation does not enter into the likelihood and is not estimated.

Comparison of Censored-Inflated, Heckman, and Two-Part

- Like the censored-inflated and Heckman models, the two-part model has different regression equations for the two parts
- Unlike the censored-inflated model, the two-part model does not have a mixture at zero, nor does Heckman
- Unlike the Heckman model, the two-part model does not estimate a residual correlation between the two parts
- Duan et al. (1983) pointed to two advantages of the two-part model over Heckman:
  - Applied to medical care expenses, it is preferable to the Heckman model because the censoring point of zero expense does not represent missing data but rather a real, observed value
  - A bivariate normality assumption for the residuals is not needed

Example: Comparing Methods on Heavy Drinking Data



NLSY Data on Heavy Drinking ($n = 1,152$)

- Dependent variable: frequency of heavy drinking measured by the question:
  - "How often have you had 6 or more drinks on one occasion during the last 30 days?"
  - Never (0); once (1); 2 or 3 times (2); 4 or 5 times (3); 6 or 7 times (4); 8 or 9 times (5); and 10 or more times (6)
- Covariates: gender, ethnicity, early onset of regular drinking (es), family history of problem drinking, and high school dropout.

Input for Censored-Normal (Tobit) and Censored-Inflated

	USEVARIABLES = hd84 male black hisp es fh123 hsdrp;
	CENSORED = hd84 (B);
ANALYSIS:	ESTIMATOR = MLR;
MODEL:	**hd84 ON male black hisp es fh123 hsdrp;**

	USEVARIABLES = hd84 male black hisp es fh123 hsdrp;
	CENSORED = hd84 (BI);
ANALYSIS:	ESTIMATOR = MLR;
MODEL:	**hd84 ON male black hisp es fh123 hsdrp;**
	hd84#1 ON male black hisp es fh123 hsdrp;

DATA TWOPART

The DATA TWOPART command is used to create a binary and a continuous variable from a continuous variable with a floor effect. A cutpoint of zero is used as the default. Following are the rules used to create the two variables:

1. If the value of the original variable is missing, both the new binary and the new continuous variable values are missing
2. If the value of the original variable is greater than the cutpoint value, the new binary variable value is one and the new continuous variable value is the log of the original variable as the default
3. If the value of the original variable is less than or equal to the cutpoint value, the new binary variable value is zero and the new continuous variable value is missing

Input for Heckman and Two-Part

	USEVARIABLES = male black hisp es fh123 hsdrp **u positive;**
	CATEGORICAL = u;
DATA TWOPART:	
	NAMES = hd84;
	BINARY = u;
	CONTINUOUS = positive;
ANALYSIS:	ESTIMATOR = MLR;
	LINK = PROBIT;
	MCONVERGENCE = 0.00001;
	INTEGRATION = 30;
MODEL:	**positive u ON male black hisp es fh123 hsdrp;**
	f BY u positive ; f@1;

	USEVARIABLES = male black hisp es fh123 hsdrp **u positive;**
	CATEGORICAL = u;
DATA TWOPART:	
	NAMES = hd84;
	BINARY = u;
	CONTINUOUS = positive;
ANALYSIS:	ESTIMATOR = MLR;
	LINK = PROBIT;
MODEL:	**positive u ON male black hisp es fh123 hsdrp;**
OUTPUT:	TECH1 TECH8;

Loglikelihood and BIC for Four Models for Frequency of Heavy Drinking

The Heckman and two-part models use $log(y)$ so logL and BIC values cannot be compared to those of tobit and censored-inflated:

Model	log L	# parameters	BIC
Censored-normal (tobit)	-1530.512	8	3117
Censored-inflated	-1499.409	15	3105
Sample selection (Heckman)	-1088.182	16	2289
Two-part	-1088.400	15	2283

Results for the censored-normal (tobit) regression model

Parameter	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
hd84 ON				
male	2.106	0.210	10.038	0.000
black	-2.157	0.258	-8.359	0.000
hisp	-1.059	0.298	-3.555	0.000
es	0.716	0.286	2.503	0.012
fh123	0.615	0.317	1.938	0.053
hsdrp	0.240	0.265	0.908	0.364
Intercepts				
hd84	-1.258	0.211	-5.961	0.000
Residual variances				
hd84	8.678	0.559	15.525	0.000

Results for the censored-inflated regression model

Parameter	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
hd84 ON				
male	0.957	0.236	4.057	0.000
black	-1.150	0.282	-4.073	0.000
hisp	-0.405	0.320	-1.264	0.206
es	0.585	0.276	2.120	0.034
fh123	-0.031	0.329	-0.095	0.924
hsdrp	0.390	0.263	1.487	0.137
hd84#1 ON				
male	-1.025	0.166	-6.157	0.000
black	0.962	0.208	4.621	0.000
hisp	0.570	0.215	2.651	0.008
es	-0.204	0.198	-1.032	0.302
fh123	-0.512	0.273	-1.876	0.061
hsdrp	0.040	0.188	0.213	0.831
Intercepts				
hd84#1	0.412	0.145	2.848	0.004
hd84	1.567	0.189	8.290	0.000

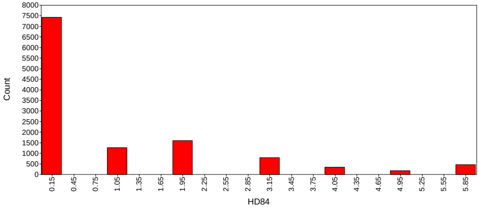
Comparisons of Results

- Heckman versus Two-part:
  - Very similar logL/BIC and results (the Heckman probit coefficients need to be divided by $\sqrt{2}$ due to adding the factor)
  - The Heckman residual correlation is significant
- Censored-inflated versus Two-part:
  - Similar results (reverse signs for the binary part)
  - LogL and BIC not comparable but limited model fit comparison can be made using MODEL CONSTRAINT:

Table : Estimated probability of zero heavy drinking and mean of heavy drinking for a subset of males who have zero values on the covariates black, hisp, es, fh123, and hsdrp

	Probability	Mean
Sample values	0.441	1.538
Censored-inflated estimates	0.402	1.547
Two-part estimates	0.403	1.671

Heckman and Two-Part Treating the Positive Part as Ordinal



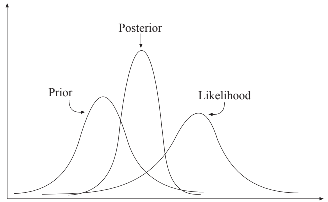
- Assignment: As an alternative, an ordinal approach may be good for these data given
  1. the limited number of response categories
  2. the slight ceiling effect for category 6, 10 or more times so that the assumption of a log normal distribution can be questioned:

  - Declare the positive part as categorical using the CATEGORICAL option of the VARIABLE command
  - Use TRANSFORM = NONE in the DATA TWOPART command to avoid the log transformation

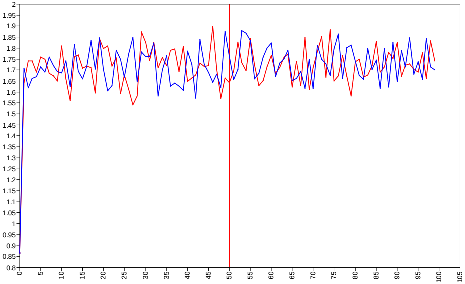
Bayesian Analysis: Advantages over ML

- Six key advantages of Bayesian analysis over frequentist analysis using maximum likelihood estimation:
  1. More can be learned about parameter estimates and model fit
  2. Small-sample performance is better and large-sample theory is not needed
  3. Parameter priors can better reflect results of previous studies
  4. Analyses are in some cases less computationally demanding, for example, when maximum-likelihood requires high-dimensional numerical integration
  5. In cases where maximum-likelihood computations are prohibitive, Bayes with non-informative priors can be viewed as a computing algorithm that would give essentially the same results as maximum-likelihood if maximum-likelihood estimation were computationally feasible
  6. New types of models can be analyzed where the maximum-likelihood approach is not practical

Figure : Prior, likelihood, and posterior for a parameter



- Priors:
  - Non-informative priors (diffuse priors): Large variance (default in Mplus)
    - A large variance reflects large uncertainty in the parameter value. As the prior variance increases, the Bayesian estimate gets closer to the maximum-likelihood estimate
  - Weakly informative priors: Used for technical assistance
  - Informative priors:
    - Informative priors reflect prior beliefs in likely parameter values
    - These beliefs may come from substantive theory combined with previous studies of similar populations



Potential scale reduction criterion (Gelman & Rubin, 1992):

$$PSR = \sqrt{\frac{W+B}{W}}, \quad (20)$$

where W represents the within-chain variation of a parameter and B represents the between-chain variation of a parameter. A PSR value close to 1 means that the between-chain variation is small relative to the within-chain variation and is considered evidence of convergence.

Markov Chain Monte Carlo (MCMC) Algorithm

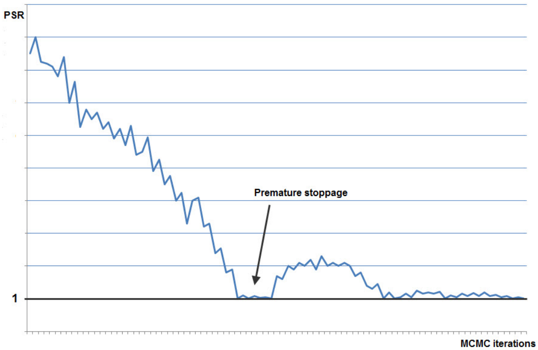
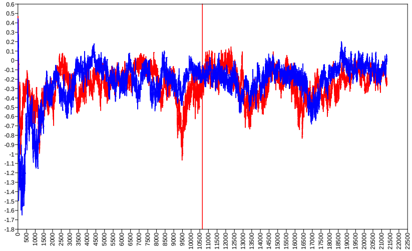
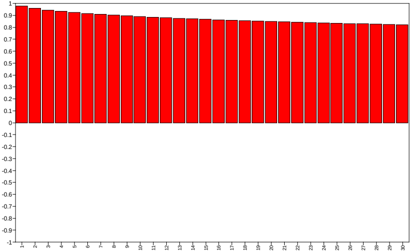


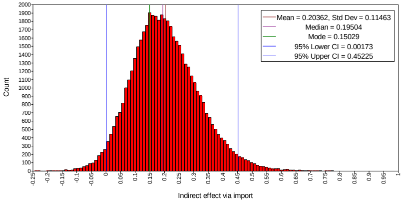
Figure : Premature stoppage of Bayes MCMC iterations using the Potential Scale Reduction (PSR) criterion

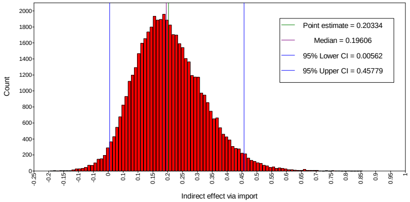
Trace and Autocorrelation Plots Indicating Poor Mixing





Bayes Posterior Distribution Similar to ML Bootstrap Distribution: Credibility versus Confidence Intervals





Bayes' Advantage Over ML: Informative Priors

- Frequentists often object to Bayes using informative priors
- But they already do use such priors in many cases in unrealistic ways (e.g. factor loadings fixed exactly at zero)
- Bayes can let informative priors reflect prior studies
- Bayes can let informative priors identify models that are unidentified by ML which is useful for model modification (BSEM)
- The credibility interval for the posterior distribution is narrower with informative priors

Speed Of Bayes In Mplus

Wang & Preacher (2014). Moderated mediation analysis using Bayesian methods. *Structural Equation Modeling*.

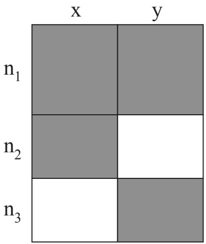
- Comparison of ML (with bootstrap) and Bayes: Similar statistical performance
- Comparison of Bayes using BUGS versus Mplus: Mplus is 15 times faster
- Reason for Bayes being faster in Mplus:
  - Mplus uses Fortran (fastest computational environment)
  - Mplus uses parallel computing so each chain is computed separately
  - Mplus uses the largest updating blocks possible - complicated to program but gives the best mixing quality
  - Mplus uses sufficient statistics
- Mplus Bayes considerably easier to use

Regressing y On x: Bringing x's Into The Model

ML estimation maximizes the log likelihood for the bivariate distribution of y and x expressed as,

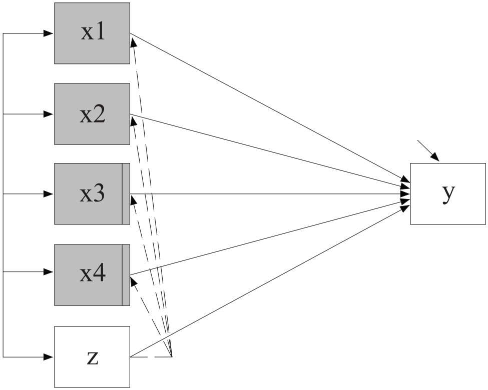
$$logL = \sum_i log[y_i, x_i] = \sum_{i=1}^{n_1} log[y_i \mid x_i] + \sum_{i=1}^{n_1+n_2} log[x_i] + \sum_{i=n_1+n_2+1}^{n_1+n_2+n_3} log[y_i].$$

Figure : Missing data patterns. White areas represent missing data



Example: Monte Carlo Simulation Study

- Linear regression with 40% missing on x_1 - x_4; no missing on y
- x_3 and x_4 s are binary split 86/16
- MAR holds as a function of the covariate z with no missing
- $n = 200$
- Comparison of Bayes and ML



```
DATA:        FILE = MARn200replist.dat;
             TYPE = MONTECARLO;
VARIABLE:    NAMES = y x1-x4 z;
             USEVARIABLES = y x1-z;
             CATEGORICAL = x3-x4;
DEFINE:      IF(z gt .25)THEN x1=_MISSING;
             IF(z gt .25)THEN x2=_MISSING;
             IF(-z gt .25)THEN x3=_MISSING;
             IF(-z gt .25)THEN x4=_MISSING;
ANALYSIS:    ESTIMATOR = BAYES;
             PROCESSORS = 2;
             BITERATIONS = (10000);
             MEDIATOR = OBSERVED;
MODEL:       y ON x1-z*.5;
             y*1;
             x1-z WITH x1-z;
```

ML Versus Bayes Treating Binary X's As Binary

- Attempting to estimate the same model using ML leads to much heavier computations due to the need for numerical integration over several dimensions
- Already in this simple model ML requires three dimensions of integration, two for the x_3, x_4 covariates and one for a factor capturing the association between x_3 and x_4.
- Bayes uses a multivariate probit model that generates correlated latent response variables underlying the binary x's - no need for numerical integration

Bayes' Advantage Over ML: Missing Data with a Binary Outcome

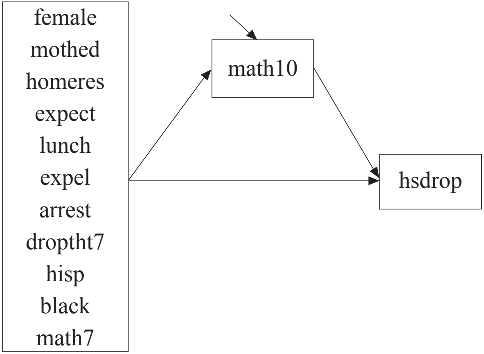


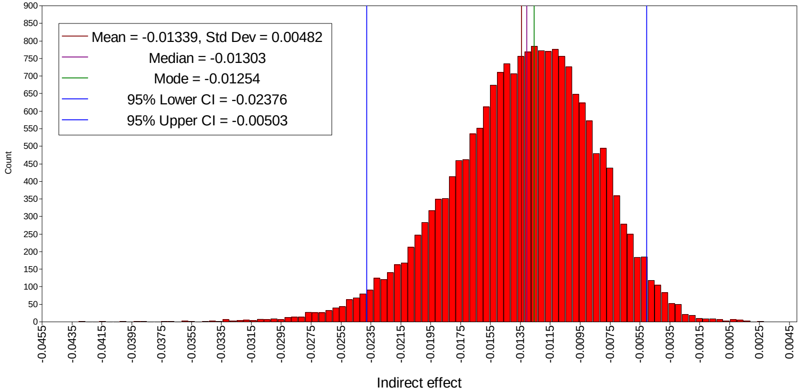
Figure : Mediation model for a binary outcome of dropping out of high school (n=2898)

Bayes With Missing Data On The Mediator

	CATEGORICAL = hsdrop;
ANALYSIS:	ESTIMATOR = BAYES;
	PROCESSORS = 2;
	BITERATIONS = (20000);
MODEL:	hsdrop ON math10 female-math7;
	math10 ON female-math7;
MODEL INDIRECT:	
	hsdrop IND math10 math7(61.01 50.88);
OUTPUT:	SAMPSTAT PATTERNS TECH1 TECH8 CINTERVAL;
PLOT:	TYPE = PLOT3;

Indirect and direct effects computed in probability scale using counterfactually-based causal effects.

Bayesian Posterior Distribution Of Indirect Effect For High School Dropout



ML estimates are almost identical to Bayes, but:

- ML needs Monte Carlo integration with 250 points because the mediator is a partially latent variable due to missing data
- ML needs bootstrapping (1,000 draws) to capture CIs for the non-normal indirect effect
- ML takes 21 minutes

- Bayes takes 21 seconds
- Bayes posterior distribution for the indirect effect is based on 20,000 draws as compared to 1,000 bootstraps for ML

Missing On The Mediator And The Covariates
Treating All Covariates As Normal: ML Versus Bayes

- ML requires integration over 10 dimensions
- ML needs 2,500 Monte Carlo integration points for sufficient precision
- ML takes 6 hours with 1,000 bootstraps

- Bayes takes less than a minute
- Bayes posterior based on 20,000 draws as compared to 1,000 bootstraps for ML

Missing On The Mediator And The Covariates
Treating Binary Covariates As Binary: ML Versus Bayes

6 covariates are binary.

- ML requires 10 + 15 = 35 dimensions of integration: intractable
- Bayes takes 3 minutes for 20,000 draws

Figure : A basic mediation model with an exposure variable *x*, a control variable *c*, a mediator *m*, and an outcome *y*

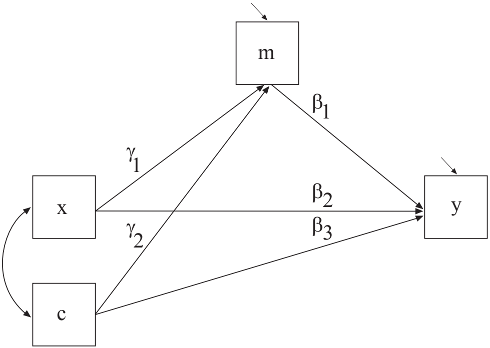
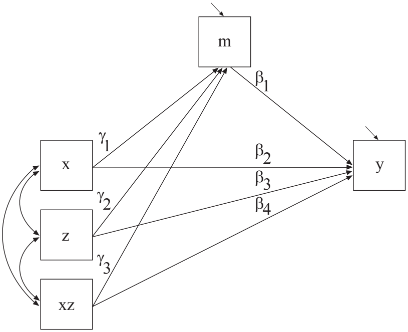


Figure : Case 1 moderated mediation of y on x, m on x, both moderated by z

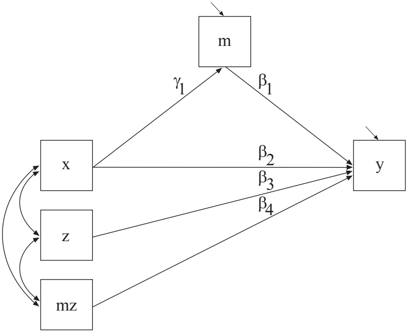


$$Indirect: \beta_1\ (\gamma_1 + \gamma_3\ z)(x_1 - x_0), \quad (21)$$

$$Direct: (\beta_2 + \beta_4\ z)(x_1 - x_0). \quad (22)$$

Moderated Mediation Analysis: Case 2 (mz)

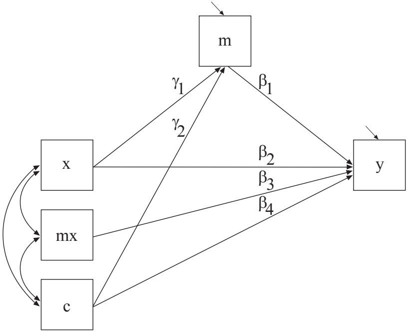
Figure : Case 2 moderated mediation of y on m moderated by z



$$Indirect : (\beta_1 + \beta_4\, z)\gamma_1(x_1 - x_0), \quad (23)$$

$$Direct : \beta_2(x_1 - x_0). \quad (24)$$

Figure : Case 3 moderated mediation of y on m moderated by x

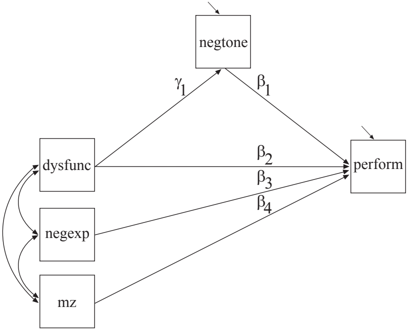


$$Indirect : (\beta_1 + \beta_3\, x_1)\gamma_1(x_1 - x_0), \tag{25}$$

$$Direct : (\beta_2 + \beta_3(\gamma_0 + \gamma_1\, x_0 + \gamma_2\, c))(x_1 - x_0). \tag{26}$$

Example: Case 2 Moderated Mediation for Work Team Performance (Hayes, 2013; $n = 60$)

Figure : Case 2 (*mz*) moderated mediation for work team behavior. The exposure variable is dysfunc (continuous). The interaction variable *mz* is the product of the mediator variable negtone and the moderator variable negexp

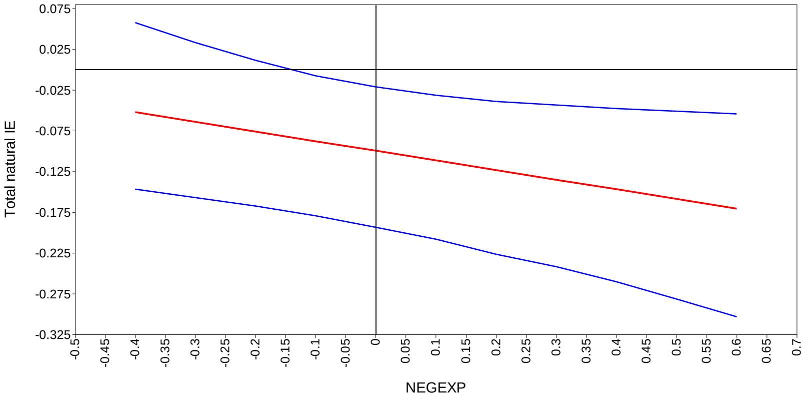


Input for Case 2 Moderated Mediation for Work Teams

TITLE:	Hayes (2013) TEAMS Case 2 moderation of M ->Y
DATA:	FILE = teams.txt;
VARIABLE:	NAMES = dysfunc negtone negexp perform;
	USEVARIABLES = dysfunc negtone negexp perform mz;
DEFINE:	**mz = negtone*negexp;**
ANALYSIS:	ESTIMATOR = ML;
	BOOTSTRAP = 10000;
MODEL:	perform ON negtone dysfunc negexp mz;
	negtone ON dysfunc;
MODEL INDIRECT:	
	perform MOD negtone negexp(-.4,.6,.1)
	mz dysfunc(0.4038 0.035);
OUTPUT:	SAMPSTAT **CINTERVAL(BOOTSTRAP);**
PLOT:	TYPE = PLOT3;

- The moderator variable negexp has 20^{th} and 80^{th} percentiles -0.4 and 0.6, respectively
- The exposure variable dysfunc has mean 0.4038 and standard deviation 0.369 so that $x_1 - x_0 = 0.4038 - 0.035 = 0.369$. In other words, 0.035 is one standard deviation below the mean

Figure : Indirect effect and bootstrap confidence interval for case 2 (*mz*) moderated mediation for work team behavior. The moderator variable is negexp and the indirect effect is labeled Total natural IE

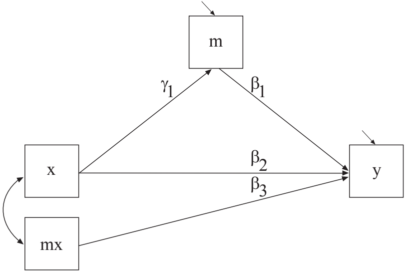


Ignore Chi-Square Test of Model Fit When Interaction Involves the Mediator

An alternative specification used in Preacher et al. (2007) avoids the two degrees of freedom that arise because of the two left-out arrows in the model. This saturates the model by allowing covariances between the moderator variable and the mediator residual and between the moderator-exposure interaction variable and the mediator residual. To accomplish this, the MODEL specification adds a line using WITH:

```
MODEL:
        perform ON negtone dysfunc negexp mz;
        negtone ON dysfunc;
        negexp mz WITH negtone dysfunc;
```

Example: Case 3 Moderated Mediation



The effects of x on y are

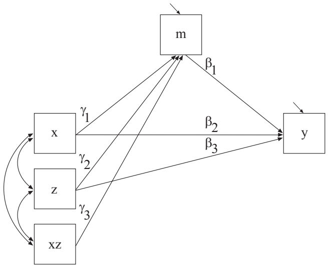
$$Indirect: (\beta_1 + \beta_3\, x_1)\gamma_1(x_1 - x_0), \quad (27)$$
$$Direct: (\beta_2 + \beta_3(\gamma_0 + \gamma_1\, x_0))(x_1 - x_0). \quad (28)$$

Non-Significant Exposure-Mediator Interaction

Quoting VanderWeele (2015, p. 46):

> *"An investigator might be tempted to only include such exposure-mediator interactions in the model if the interaction is statistically significant. - - This approach is problematic. It is problematic because power to detect interaction tends to be very low unless the sample size is very large. - - such exposure-mediator interaction may be important in capturing the dynamics of mediation... - - A better approach - - is perhaps to include them by default and only exclude them if they do not seem to change the estimates of the direct and indirect effects very much."*

TITLE:	x moderation of y regressed on m
DATA:	FILE = xmVx4s1n200rep6.dat;
VARIABLE:	NAMES = y m x;
	USEVARIABLES = y m x mx;
DEFINE:	**mx = m*x;**
ANALYSIS:	ESTIMATOR = ML;
	BOOTSTRAP = 10000;
MODEL:	**y ON m x mx;**
	m ON x;
MODEL INDIRECT:	
	y MOD m mx x(7 5);
OUTPUT:	SAMPSTAT CINTERVAL(BOOTSTRAP);
PLOT:	TYPE = PLOT3;



The model used for data generation is

$$y_i = \beta_0 + \beta_1\, m_i + \beta_2\, x_i + \beta_3\, z_i + \varepsilon_{yi}, \tag{29}$$

$$m_i = \gamma_0 + \gamma_{1i}\, x_i + \gamma_2\, z_i + \varepsilon_{mi}, \tag{30}$$

$$\gamma_{1i} = \gamma_1 + \gamma_3\, z_i, \tag{31}$$

where γ_{1i} is a random slope. Inserting (31) in (30) shows that the random slope formulation is equivalent to adding an interaction term xz as a covariate in the regression of m.

Input for Simulation of *z* Moderation of *m* Regressed on *x*

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 = xzrep*.dat;
	CUTPOINTS = x(0);
MODEL POPULATION:	
	x-z@1; [x-z@0];
	x WITH z@0.5;
	y ON m*.5 x*.2 z*.1; y*.5; [y*0];
	gamma1 | m ON x;
	[gamma1*.3];
	gamma1 ON z*.2;
	gamma1@0;
	m ON z*.3; m*1; [m*0];
ANALYSIS:	**TYPE = RANDOM;**
MODEL:	y ON m*.5 (b)
	x*.2 z*.1;
	y*.5; [y*0];
	gamma1 | m ON x;
	[gamma1*.3] (gamma1);
	gamma1 ON z*.2 (gamma3);
	gamma1@0;
	m ON z*.3; m*1; [m*0];
MODEL CONSTRAINT:	
	NEW(indavg*.15 indlow*.05 indhigh*.25);
	indavg = b*gamma1;
	indlow = b*(gamma1-gamma3);
	indhigh = b*(gamma1+gamma3);

Results for Monte Carlo Simulation of z Moderation of m Regressed on x using $n = 400$ and 500 Replications

	Population	Average	Std. Dev.	S.E. Average	M.S.E.	95% Cover	% Sig Coeff
gamma1 ON							
z	0.200	0.2010	0.0775	0.0771	0.0060	0.950	0.744
y ON							
m	0.500	0.5007	0.0524	0.0494	0.0027	0.922	1.000
x	0.200	0.2056	0.0783	0.0784	0.0061	0.938	0.754
z	0.100	0.0963	0.0470	0.0433	0.0022	0.926	0.604
m ON							
z	0.300	0.2999	0.0531	0.0545	0.0028	0.964	1.000
Intercepts							
y	0.000	-0.0017	0.0527	0.0522	0.0028	0.934	0.066
m	0.000	-0.0008	0.0543	0.0545	0.0029	0.946	0.054
gamma1	0.300	0.3010	0.0776	0.0770	0.0060	0.962	0.978
Residual Variances							
y	0.500	0.4938	0.0341	0.0347	0.0012	0.928	1.000
m	0.500	0.4940	0.0331	0.0346	0.0011	0.950	1.000
gamma1	0.000	0.0000	0.0000	0.0000	0.0000	1.000	0.000
New/Additional Parameters							
indavg	0.150	0.1505	0.0417	0.0416	0.0017	0.956	0.974
indlow	0.050	0.0497	0.0546	0.0548	0.0030	0.958	0.138
indhigh	0.250	0.2514	0.0628	0.0603	0.0039	0.928	0.988

Figure : Mediator-outcome confounding 1

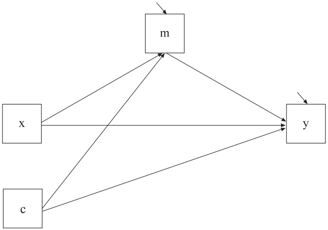
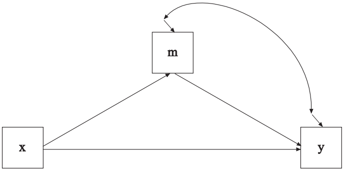
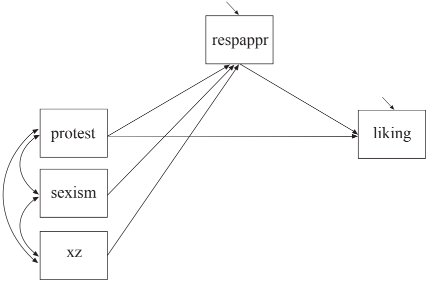


Figure : Mediator-outcome confounding 2



Sensitivity Analysis for Discrimination Study (Hayes, 2013)



A moderated mediation model of sex discrimination in the work place. The interaction variable xz is the product of the exposure variable protest and the moderator variable sexism ($n = 129$)

- Variables:
  - Protest: binary exposure variable (2 randomized scenarios of female attorney taking action or not)
  - Sexism: Moderator variable
  - Respappr: Mediator - perceived appropriateness of response)
  - Liking: Outcome - how well the subject likes the female attorney

Results for Combined Moderated Mediation for Sex Discrimination

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
liking ON				
respappr	0.098	0.533	0.184	0.854
protest	-3.119	1.750	-1.782	0.075
sexism	-0.462	0.502	-0.919	0.358
mx	0.112	0.157	0.715	0.475
mz	0.039	0.100	0.392	0.695
xz	0.500	0.341	1.466	0.143
respappr ON				
protest	-2.687	1.738	-1.546	0.122
sexism	-0.529	0.320	-1.654	0.098
xz	0.810	0.346	2.343	0.019
Intercepts				
liking	6.510	2.623	2.482	0.013
respappr	6.567	1.596	4.114	0.000
Residual Variances				
liking	0.779	0.135	5.767	0.000
respappr	1.269	0.156	8.121	0.000

Figure : Loop plot of indirect effect and confidence interval for combined moderated mediation case of sex discrimination. The moderator is labeled z in MODEL CONSTRAINT and corresponds to the sexism variable

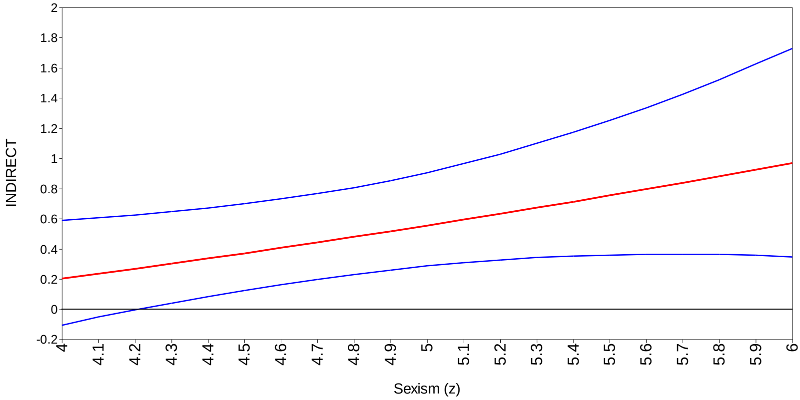
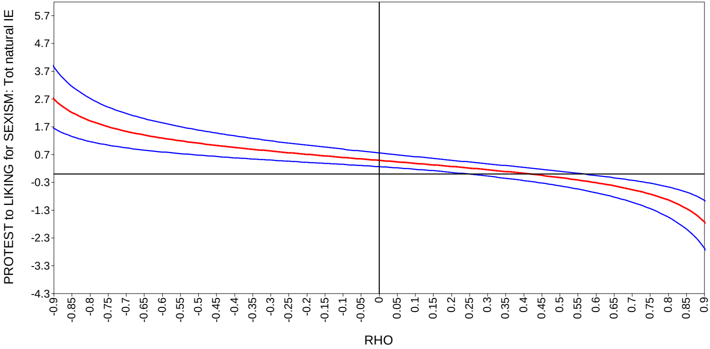


Table : Input for moderated mediation for sex discrimination data

TITLE:	Hayes PROTEST moderation of X ->M, X->Y
DATA:	FILE = protest.txt;
VARIABLE:	NAMES = sexism liking respappr protest;
	USEVARIABLES = liking respappr protest sexism xz;
DEFINE:	xz = protest*sexism;
ANALYSIS:	ESTIMATOR = ML;
	BOOTSTRAP = 1000;
MODEL:	liking ON respappr (beta1)
	protest (beta2)
	sexism
	xz (beta4);
	respappr ON protest (gamma1)
	sexism (gamma2)
	xz (gamma3);
MODEL INDIRECT:	
	liking MOD respappr sexism(4,6,.1) xz protest;
OUTPUT:	SAMPSTAT STANDARDIZED
	CINTERVAL(BOOTSTRAP);
PLOT:	**TYPE = PLOT3 SENSITIVITY;**

Figure : Sensitivity plot for the indirect effect and its confidence interval at the sexism mean of 5 in a study of sex discrimination in the workplace. The x-axis represents the residual correlation ρ and the y-axis represents the indirect effect



Counterfactually-Defined Causal Effects: Potential Outcomes, Counterfactuals, and Causal Effects

i	X_i	Potential Outcomes $Y_i\ (X_i{=}1)$	Potential Outcomes $Y_i\ (X_i{=}0)$	Causal effect $Y_i\ (X_i{=}1) - Y_i\ (X_i{=}0)$
1	1	[11]	9	2
2	1	[14]	10	4
3	0	8	[5]	3
4	1	[9]	8	1
5	0	18	[12]	6
6	0	11	[10]	1
True average		11.83	9	2.83
Observed average		11.33	9	2.33

Counterfactually-Defined Causal Effects: Robins, Pearl, VanderWeele, Imai

- Counterfactuals and potential outcomes:
  - Chapter 4: continuous mediator and continuous outcome
  - Chapter 8: continuous mediator and binary outcome, binary mediator and continuous or binary outcome, count outcome, two-part outcome
- Counterfactually-defined causal indirect and direct effects:
  - Strict assumptions including no mediator-outcome confounding
  - X = exposure variable, M = mediator, Y = outcome
  - Total effect: $E[Y(1,M(1))] - E[Y(0,M(0))]$, treatment group mean of Y minus control group mean of Y
  - The Total Natural Indirect Effect (TNIE) $= E[Y(1,M(1))] - E[Y(1,M(0))]$ where 1 and 0 represent treatment and control for the exposure variable
  - What does it mean?
  - Explanations in words and formulas

Indirect Effect $TNIE = E[Y(1,M(1))] - E[Y(1,M(0))]$

- In words:
  - $E[Y(1,M(1))]$ is the mean of the outcome when subjects get the treatment ($X = 1$) and M varies as it would under the treatment condition ($X = 1$) - this is the treatment group mean
  - $E[Y(1,M(0))]$ is the mean of the outcome when subjects get the treatment ($X = 1$) but M varies as it would under the control condition ($X = 0$) - this is a counterfactual
- In formulas:
  - To get an effect of X on Y we need to integrate out M
  - M has two different distributions $f(M|X)$: $M(0)$ for $X = 0$ and $M(1)$ for $X = 1$. For example:
  - $E[Y(1,M(0))] = \int_{-\infty}^{+\infty} E[Y|X = 1, M = m] \times f(M|X = 0)\, \partial M$
  - In some cases, this integral is simple - integration does not need to be involved: (1) Continuous M, continuous Y, (2) Continuous M, binary Y with probit
  - In some cases, the integration is needed: (1) Continuous M, binary Y with logistic (numerical integration needed), (2) Count Y, (3) $log(Y)$

- Continuous M and Y:

$$Y_i = \beta_0 + \beta_1 \, M_i + \beta_2 \, X_i + \varepsilon_{yi}, \tag{32}$$

$$M_i = \gamma_0 + \gamma_1 \, X_i + \varepsilon_{mi}. \tag{33}$$

Inserting (33) in (32) and integrating over M,

$$\begin{aligned} E[Y(x_1, M(x_0))] &= \beta_0 + \beta_2 \, x_1 + \\ &+ \beta_1 \int_{-\infty}^{+\infty} M \, f(M; \gamma_0 + \gamma_1 \, x_0, \sigma^2) \, \partial M, \\ &= \beta_0 + \beta_2 \, x_1 + \beta_1 (\gamma_0 + \gamma_1 \, x_0). \end{aligned} \tag{34}$$

Conditioning on $X = x_1$ in (32) and $X = x_0$ in (33) and inserting the mediator expression in the outcome expression, the expected value is the same:

$$= \beta_0 + \beta_2 \, x_1 + \beta_1 (\gamma_0 + \gamma_1 \, x_0). \tag{35}$$

- $TNIE$ for continuous M and Y:

$$E[Y(x_1, M(x_1))] - E[Y(x_1, M(x_0))] \tag{36}$$
$$= \beta_0 + \beta_2\, x_1 + \beta_1(\gamma_0 + \gamma_1\, x_1) \tag{37}$$
$$- (\beta_0 + \beta_2\, x_1 + \beta_1(\gamma_0 + \gamma_1\, x_0)) \tag{38}$$
$$= \beta_1\, \gamma_1(x_1 - x_0). \tag{39}$$

- Note 1: Often $x_1 - x_0 = 1$ such as with a one-unit change or treatment/control.
- Note 2: β_0, γ_0, β_2 cancel out. The indirect effect is a product of 2 slopes. This is not the case for binary Y

Now We Know How To Do *TNIE* for Binary *Y*

$$Y_i^* = \beta_0 + \beta_1\, M_i + \beta_2\, X_i + \varepsilon_{yi}, \quad (40)$$

$$M_i = \gamma_0 + \gamma_1\, X_i + \varepsilon_{mi}. \quad (41)$$

Conditioning on $X = x_1$ and $X = x_0$, for Y^* and M, respectively, and inserting M into Y,

$$E(Y^*|X) = \beta_0 + \beta_1\, \gamma_0 + \beta_1\, \gamma_1\, x_0 + \beta_2\, x_1, \quad (42)$$

$$V(Y^*|X) = V(\beta_1\, \varepsilon_m + \varepsilon_y) = \beta_1^2\, \sigma_m^2 + c. \quad (43)$$

$$P(Y = 1|X) = \Phi[E(Y^*|X)/\sqrt{V(Y^*|X)}], \quad (44)$$

$$TNIE = \Phi[1,1] - \Phi[1,0], \quad (45)$$

where $\Phi[1,1]$ uses $\beta_0 + \beta_1\, \gamma_0 + \beta_1\, \gamma_1\, x_1 + \beta_2\, x_1$ in $E(Y^*|X)$ and $\Phi[1,0]$ uses $\beta_0 + \beta_1\, \gamma_0 + \beta_1\, \gamma_1\, x_0 + \beta_2\, x_1$. All 6 parameters involved.

Effects Expressed on an Odds Ratio Scale for a Binary Outcome: Probit Model

The total natural indirect effect odds ratio for a binary exposure can be expressed as

$$TNIE(OR) = \frac{P(Y_{x_1 M_{x_1}} = 1)/(1 - P(Y_{x_1 M_{x_1}} = 1)}{P(Y_{x_1 M_{x_0}} = 1)/(1 - P(Y_{x_1 M_{x_0}} = 1))}$$
$$= \frac{\Phi[probit(1,1)]/(1 - \Phi[probit(1,1)])}{\Phi[probit(1,0)]/(1 - \Phi[probit(1,0)])}. \qquad (46)$$

Odds Ratio Effects Assuming a Rare Binary Outcome: Logistic Model

VanderWeele and Vansteelandt (2010) show that with logistic regression the TNIE odds ratio is approximately equal to

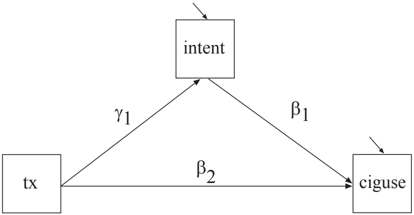
$$TNIE(OR) \approx e^{\beta_1\,\gamma_1 + \beta_3\,\gamma_1}, \tag{47}$$

that is, the indirect effect odds ratio uses the same formula as the indirect effect with a continuous outcome, but exponentiated. When the treatment variable is continuous, the indirect effect odds ratio of (47) is modified as

$$TNIE(OR) = e^{(\beta_1\,\gamma_1 + \beta_3\,\gamma_1\,x_1)(x_1 - x_0)}, \tag{48}$$

for a change from x_0 to x_1. For example, x_0 may represent the mean of the treatment and x_1 may represent the mean plus one standard deviation, so that $x_1 - x_0$ corresponds to one standard deviation for the continuous treatment variable.

Example: Smoking Data



Drug intervention program for students in Grade 6 and Grade 7 in Kansas City schools ($n = 864$). MacKinnon et al. (2007), Clinical Trials.

- Schools were randomly assigned to the treatment or control group (the multilevel aspect of the data is ignored)
- The mediator is the intention to use cigarettes in the following 2-month period which was measured about six months after baseline
- The outcome is cigarette use or not in the previous month which was measured at follow-up
- Cigarette use is observed for 18% of the sample

- The total effect can be computed without doing a mediation analysis as the difference between the proportion of smokers in the treatment group and the proportion of smokers in the control group
- This results in an estimate of the total effect as the difference in the probabilities of $0.148 - 0.224 = -0.076$
- The corresponding estimate of the total effect odds ratio is

$$TE(OR) = \frac{0.148/(1-0.148)}{0.224/(1-0.224)} = 0.602. \quad (49)$$

- Both estimates indicate a lowering of the smoking probability due to treatment

Table : Input for smoking data using probit

TITLE:	Clinical Trials data from MacKinnon et al. (2007)
DATA:	FILE = smoking.txt;
VARIABLE:	NAMES = intent tx ciguse;
	USEVARIABLES = tx ciguse intent;
	CATEGORICAL = ciguse;
ANALYSIS:	**ESTIMATOR = ML;**
	LINK = PROBIT;
	BOOTSTRAP = 10000;
MODEL:	ciguse ON intent tx;
	intent ON tx;
MODEL INDIRECT:	
	ciguse IND intent tx;
OUTPUT:	TECH1 TECH8 SAMPSTAT
	CINTERVAL(BOOTSTRAP);
PLOT:	TYPE = PLOT3;

Table : Bootstrap confidence intervals for smoking data effects using probit regression for the outcome cigarette

Confidence intervals of total, indirect, and direct effects based on counterfactuals (causally-defined effects)					
	Lower 2.5%	Lower 5%	Estimate	Upper 5%	Upper 2.5%
Effects from tx to ciguse					
Tot natural IE	-0.040	-0.036	-0.022	-0.008	-0.006
Pure natural DE	-0.104	-0.095	-0.050	-0.005	0.004
Total effect	-0.128	-0.119	-0.072	-0.026	-0.017
Odds ratios for binary Y					
Tot natural IE	0.757	0.772	0.853	0.939	0.958
Pure natural DE	0.520	0.551	0.731	0.969	1.025
Total effect	0.433	0.461	0.624	0.841	0.896

- The total natural indirect effect (TNIE) in probability metric is estimated as -0.022 and is significant because the 95% confidence interval does not cover zero: $[-0.040, -0.006]$
- The indirect effect odds ratio is estimated as 0.853 and is significant because the 95% confidence interval does not cover one: $[0.757, 0.958]$
- The direct effect in probability metric is estimated as -0.050 and is not significant. The direct effect odds ratio of 0.731 is not significant
- The total effect in probability metric of -0.072 is significant
- The total effect can be compared to the proportion of cigarette users in the control group of 0.224. This shows a drop of 34% due to treatment

Table : Input for smoking data using logistic regression for the cigarette use outcome

TITLE:	Clinical Trials data from MacKinnon et al. (2007)
DATA:	FILE = smoking.txt;
VARIABLE:	NAMES = intent tx ciguse;
	USEVARIABLES = tx ciguse intent;
	CATEGORICAL = ciguse;
ANALYSIS:	ESTIMATOR = ML;
	LINK = LOGIT;
	BOOTSTRAP = 10000;
MODEL:	ciguse ON intent (beta1)
	tx (beta2);
	intent ON tx (gamma);
MODEL INDIRECT:	
	ciguse IND intent tx;
MODEL CONSTRAINT:	
	NEW(indirect direct);
	indirect = EXP(beta1*gamma);
	direct = EXP(beta2);
OUTPUT:	TECH1 TECH8 SAMPSTAT
	CINTERVAL(BOOTSTRAP);
PLOT:	TYPE = PLOT3;

Indirect and Direct Effects for Smoking Data Using Logistic

- Not assuming a rare outcome (using MODEL INDIRECT):
  TNIE (OR) $= 0.858$, TNDE (OR) $= 0.716$
- Assuming a rare outcome (using MODEL CONSTRAINT):
  TNIE (OR) $= 0.843$, TNDE (OR) $= 0.686$

- The rare outcome results indicate stronger effects with estimates farther from one
- The rare outcome assumption may not be suitable here with 18% smoking prevalence

- Probit and logistic give similar results

Moderated Mediation with a Binary Outcome: Vaccination

- Hopfer (2012) analyzed data from a randomized control trial aimed at increasing the vaccination rate for the human papillomavirus (HPV) among college women ($n = 394$)
  - Subjects were randomized into three different intervention groups and a control group where the groups were presented with different forms of video with vaccine decision narratives
  - The mediator measures intent to get vaccinated
  - Control variables are HPV communication with parents (yes/no), age, sexually active (yes/no), and HPV knowledge
  - Only the effects of the combined peer-expert intervention are considered ($tx2$)
  - In this group, to which 25% of the sample was randomized, the vaccination rate is 22.2% whereas in the control group it is 12.0%
  - This gives an estimate of the total intervention effect in the probability metric of 0.10 and in the odds ratio metric of 2.70

Figure : Moderated mediation model for the HPV vaccination data using a logistic regression for the vaccination outcome

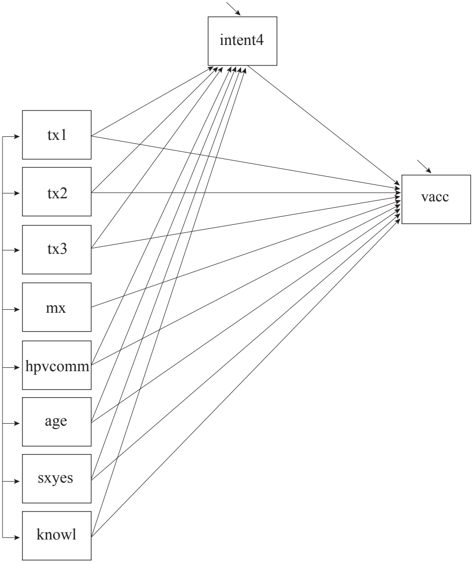


Table : Input for the model with intervention-mediator interaction for HPV vaccination data

VARIABLE:	
	USEVARIABLES = intent4 tx1 tx2 tx3 vacc hpvcomm age sxyes knowl mx;
	CATEGORICAL = vacc;
	MISSING = ALL (99);
DEFINE:	**mx = intent4*tx2;**
	CENTER age knowl(GRANDMEAN);
ANALYSIS:	ESTIMATOR = ML;
	BOOTSTRAP = 10000;
MODEL:	vacc ON intent4 tx1 tx2 tx3 hpvcomm age sxyes knowl mx;
	intent4 ON tx1 tx2 tx3 hpvcomm age sxyes knowl;
MODEL INDIRECT:	
	vacc MOD intent4 mx tx2;
OUTPUT:	SAMPSTAT PATTERNS CINTERVAL(BOOTSTRAP) TECH1 TECH8;
PLOT:	TYPE = PLOT3;

Table : Results for HPV vaccination data

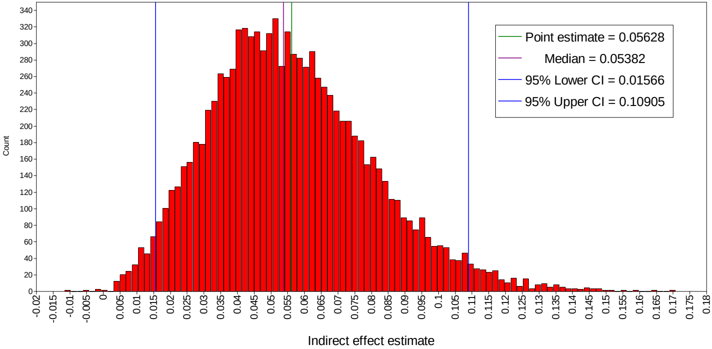
	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
vacc ON				
intent4	1.303	0.262	4.974	0.000
tx1	0.320	0.435	0.735	0.463
tx2	-1.180	2.271	-0.520	0.603
tx3	-0.818	2.141	-0.382	0.703
hpvcomm	0.242	0.350	0.693	0.488
age	0.194	0.084	2.311	0.021
sxyes	0.219	0.333	0.658	0.511
knowl	-0.041	0.072	-0.572	0.568
mx	0.494	0.660	0.749	0.454
intent4 ON				
tx1	0.149	0.106	1.400	0.161
tx2	0.300	0.092	3.270	0.001
tx3	-0.066	0.141	-0.465	0.642
hpvcomm	0.093	0.078	1.196	0.232
age	-0.049	0.021	-2.283	0.022
sxyes	0.044	0.078	0.573	0.567
knowl	-0.003	0.017	-0.160	0.873
Intercepts				
intent4	2.718	0.082	32.959	0.000
Thresholds				
vacc$1	6.227	0.877	7.100	0.000
Residual Variances				
intent4	0.591	0.041	14.293	0.000

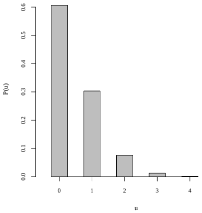
Table : Bootstrap confidence intervals without and with intervention-mediator interaction for HPV vaccination data

Confidence intervals of total, indirect, and direct effects based on counterfactuals (causally-defined effects)

	Lower 2.5%	Lower 5%	Estimate	Upper 5%	Upper 2.5%
Without intervention-mediator interaction					
Effects from TX2 to VACC					
Tot natural IE	0.016	0.020	0.048	0.083	0.092
Pure natural DE	-0.019	-0.010	0.041	0.098	0.111
Total effect	0.013	0.024	0.089	0.165	0.182
Odds ratios for binary Y					
Tot natural IE	1.155	1.197	1.448	1.833	1.932
Pure natural DE	0.803	0.894	1.523	2.715	3.045
Total effect	1.137	1.283	2.205	4.115	4.665
With intervention-mediator interaction					
Effects from TX2 to VACC					
Tot natural IE	0.016	0.020	0.056	0.099	0.109
Pure natural DE	-0.022	-0.012	0.037	0.095	0.107
Total effect	0.016	0.028	0.093	0.169	0.186
Odds ratios for binary Y					
Tot natural IE	1.147	1.200	1.541	2.096	2.238
Pure natural DE	0.773	0.865	1.467	2.662	2.964
Total effect	1.178	1.313	2.260	4.234	4.791

Figure : Bootstrap distribution for the total natural indirect effect estimate in probability metric for the model with intervention-mediator interaction for the HPV vaccination data





$$log \mu_i = \beta_0 + \beta_1 \, M_i + \beta_2 \, X_i + \beta_3 \, MX_i + \beta_4 \, C_i, \tag{50}$$

$$M_i = \gamma_0 + \gamma_1 \, X_i + \gamma_2 \, C_i + \varepsilon_{mi}. \tag{51}$$

As before, the counterfactually-based causal effects consider terms such as

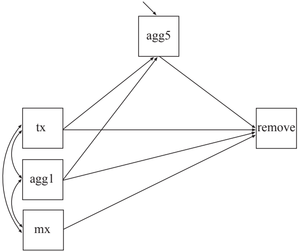
$$E[Y(x_1, M(x_0))] = \int_{-\infty}^{\infty} E[Y \mid C = c, X = x_1, M = m] \tag{52}$$

$$\times f(M \mid C = c, X = x_0) \, \partial M. \tag{53}$$

This needs to take into account that the rate (mean) is

$$E[Y \mid C = c, X = x_1, M = m] = e^{\beta_0 + \beta_1 \, m + \beta_2 \, x_1 + \beta_3 \, m \, x_1 + \beta_4 \, c}. \tag{54}$$

Example: A Mediation Model for Aggressive Behavior and a School Removal Count Outcome: Case 3 (*mx*) Moderation



Randomized field experiment in Baltimore public schools with a classroom-based intervention aimed at reducing aggressive-disruptive behavior among elementary school students (Kellam et al., 2008). The analysis uses $n = 250$ boys.

- The outcome variable remove is the number of times a student has been removed from school during grades 1-7
- tx is the binary exposure variable representing the intervention
- The Fall baseline aggression score is agg1 which was observed before the intervention started
- The mediator variable agg5 is the Grade 5 aggression score.
- An intervention-mediator interaction variable *mx* is included to moderate the influence of the mediator on the outcome.

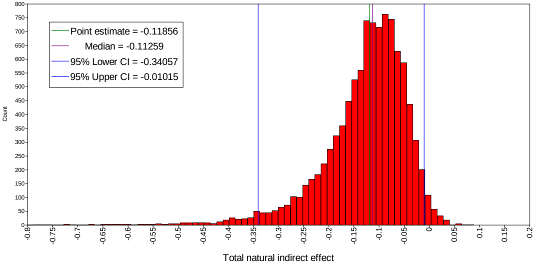
Table : Input for negative binomial model for school removal data

VARIABLE:	
	USEVARIABLES = remove agg5 agg1 tx mx;
	IDVARIABLE = prcid;
	COUNT = remove(NB);
	USEOBSERVATIONS = gender EQ 1 AND (desgn11s EQ 1 OR desgn11s EQ 2 OR desgn11s EQ 3 OR desgn11s EQ 4);
DEFINE:	IF(desgn11s EQ 4)THEN tx=1;
	IF(desgn11s EQ 1 OR desgn11s EQ 2 OR desgn11s EQ 3)THEN tx=0;
	remove = total17;
	agg1 = sctaa11f;
	agg5 = sctaa15s;
	CENTER agg1 agg5(GRANDMEAN);
	mx = agg5*tx;
ANALYSIS:	ESTIMATOR = ML;
	BOOTSTRAP = 10000;
	PROCESSORS = 8;
MODEL:	remove ON agg5 tx mx agg1;
	agg5 ON tx agg1;
MODEL INDIRECT:	
	remove MOD agg5 mx tx;
OUTPUT:	SAMPSTAT TECH1 TECH8 PATTERNS CINTERVAL(BOOTSTRAP);
PLOT:	TYPE = PLOT3;

Table : Bootstrap confidence intervals for effects for school removal data

Confidence intervals of total, indirect, and direct effects based on counterfactuals (causally-defined effects)					
	Lower 2.5%	Lower 5%	Estimate	Upper 5%	Upper 2.5%
Effects from TX to REMOVE					
Tot natural IE	-0.341	-0.283	-0.119	-0.024	-0.010
Pure natural DE	-0.681	-0.608	-0.272	0.125	0.213
Total effect	-0.794	-0.722	-0.391	-0.032	0.034
Other effects					
Pure natural IE	-0.358	-0.327	-0.183	-0.050	-0.023
Tot natural DE	-0.587	-0.525	-0.208	0.135	0.213
Total effect	-0.794	-0.722	-0.391	-0.032	0.034

Figure : Total natural indirect effect bootstrap distribution for school removal data



- The indirect effect estimate -0.119 is in a log rate metric for the count outcome of school removal and is hard to interpret
    - One way to make the effect size understandable is to compute the probability of a zero count
    - The intervention increases the probability of a zero school removals from 0.294 to 0.435

- Example from Hayes (2013):
  - $n = 262$ small-business owners' economic stress (Pollack et al., 2011)
  - The exposure variable is a continuous variable representing economic stress
  - The mediator variable is a continuous variable representing depressed affect
  - The outcome variable is a continuous variable representing thoughts about withdrawing from their entrepreneurship

The outcome variable withdraw has a 30% floor effect:

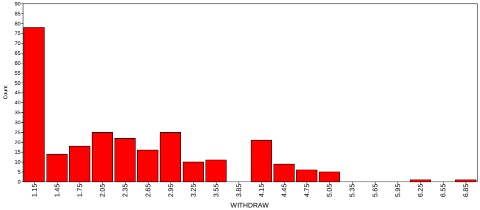


Table : Input for two-part mediation modeling of economic stress data

TITLE:	Hayes ESTRESS example, cont's X
DATA:	FILE = estress.txt;
VARIABLE:	NAMES = tenure estress affect withdraw sex age ese;
	USEVARIABLES = affect estress u y;
	CATEGORICAL = u;
DEFINE:	**withdraw = withdraw - 1;**
DATA TWOPART:	
	NAMES = withdraw;
	BINARY = u;
	CONTINUOUS = y;
	CUTPOINT = 0;
ANALYSIS:	ESTIMATOR = ML;
	LINK = PROBIT;
	BOOTSTRAP = 1000;
MODEL:	y ON affect (beta1)
	estress (beta2);
	[y] (beta0);
	y (v);
	affect ON estress (gamma1);
	[affect] (gamma0);
	affect (sig);
	u ON affect (kappa1)
	estress (kappa2);
	[u$1] (kappa0);
MODEL INDIRECT:	
	u IND affect estress (6.04 4.62);
	-table continues-

Table : Input for two-part mediation modeling of economic stress data

```
MODEL CONSTRAINT:
                NEW(x1 x0 ey1 ey0 mum1 mum0 ay1 ay0 bym11 bym10 bym01
                bym00 eym11 eym10 eym01 eym00 tnie pnde total pnie beta3 sd pi11
                pi10 pi01 pi00);
                beta3 = 0;
                x1=6.04;
                x0=4.62;
                ey1=EXP(v/2)*EXP(beta0+beta2*x1);
                ey0=EXP(v/2)*EXP(beta0+beta2*x0);
                mum1=gamma0+gamma1*x1;
                mum0=gamma0+gamma1*x0;
                ay1=sig*(beta1+beta3*x1);
                ay0=sig*(beta1+beta3*x0);
                bym11=(ay1/mum1+1);
                bym10=(ay1/mum0+1);
                bym01=(ay0/mum1+1);
                bym00=(ay0/mum0+1);
                sd=SQRT(kappa1*kappa1*sig+1);
                pi11=PHI((-kappa0+kappa2*x1+kappa1*bym11*
                (gamma0+gamma1*x1))/sd);
                pi10=PHI((-kappa0+kappa2*x1+kappa1*bym10*
                (gamma0+gamma1*x0))/sd);
                pi01=PHI((-kappa0+kappa2*x0+kappa1*bym11*
                (gamma0+gamma1*x1))/sd);
                pi00=PHI((-kappa0+kappa2*x0+kappa1*bym00*
                (gamma0+gamma1*x0))/sd);
                eym11=EXP((bym11*bym11-1)*mum1*mum1/(2*sig));
                eym10=EXP((bym10*bym10-1)*mum0*mum0/(2*sig));
                eym01=EXP((bym01*bym01-1)*mum1*mum1/(2*sig));
                eym00=EXP((bym00*bym00-1)*mum0*mum0/(2*sig));
                tnie=pi11*ey1*eym11-pi10*ey1*eym10;
                pnde=pi10*ey1*eym10-pi00*ey0*eym00;
                total=pi11*ey1*eym11-pi00*ey0*eym00;
                pnie=pi01*ey0*eym01-pi00*ey0*eym00;
PLOT:           TYPE = PLOT3;
OUTPUT:         SAMPSTAT TECH1 TECH8
                CINTERVAL(BOOTSTRAP);
```

Table : Bootstrap confidence intervals for four mediation models

Confidence intervals for effects					
	Lower 2.5%	Lower 5%	Estimate	Upper 5%	Upper 2.5%
(1) Two-part: overall effects for continuous part of the outcome					
TNIE	0.104	0.121	0.203	0.293	0.311
PNDE	-0.304	-0.276	-0.145	-0.011	0.019
TE	-0.124	-0.089	0.058	0.207	0.246
(2) Two-part: effects for binary part of the outcome					
TNIE	0.036	0.041	0.071	0.103	0.108
PNDE	-0.074	-0.062	-0.016	0.028	0.035
TE	-0.006	0.005	0.055	0.098	0.105
(3) Two-part: conditional effects for continuous part of the outcome					
TNIE	0.043	0.053	0.112	0.177	0.194
PNDE	-0.322	-0.299	-0.160	-0.008	0.023
TE	-0.219	-0.184	-0.048	0.105	0.131
(4) Regular: effects using log y					
TNIE	0.098	0.108	0.182	0.267	0.284
PNDE	-0.236	-0.209	-0.084	0.044	0.066
TE	-0.072	-0.045	0.099	0.243	0.269
(5) Regular: effects using the original y					
TNIE	0.103	0.117	0.189	0.266	0.282
PNDE	-0.263	-0.243	-0.109	0.027	0.051
TE	-0.116	-0.069	0.080	0.220	0.245