mediation: R Package for Causal Mediation Analysis

Dustin Tingley
Harvard

Teppei Yamamoto
MIT

Luke Keele
Penn State

Kosuke Imai
Princeton

Abstract

In this paper, we describe the R package **mediation** for conducting causal mediation analysis in applied empirical research. In many scientific disciplines, the goal of researchers is not only estimating causal effects of a treatment but also understanding the process in which the treatment causally affects the outcome. Causal mediation analysis is frequently used to assess potential causal mechanisms. The **mediation** package implements a comprehensive suite of statistical tools for conducting such an analysis. The package is organized into two distinct approaches. Using the model-based approach, researchers can estimate causal mediation effects and conduct sensitivity analysis under the standard research design. Furthermore, the design-based approach provides several analysis tools that are applicable under different experimental designs. This approach requires weaker assumptions than the model-based approach. Finally, we also implement a statistical method for dealing with multiple (causally dependent) mediators, which are often encountered in practice.

Keywords: causal mechanisms, mediation analysis, **mediation**, R.

1. Introduction

Scholars across a range of disciplines are increasingly interested in identifying causal mechanisms, going beyond the estimation of causal effects. Once they ascertain that certain variables causally affect the outcome, the next natural step is to understand how these variables exert their influence. The standard procedure for analyzing causal mechanisms in applied research is called *mediation analysis*, where a set of linear regression models are fitted and then the estimates of "mediation effects" are computed from the fitted models (e.g., Haavelmo 1943; Baron and Kenny 1986; Shadish, Cook, and Campbell 2001; MacKinnon 2008). In recent years, however, causal mechanisms have been studied within the modern framework of causal inference with an emphasis on the assumptions required for identification. This approach has highlighted limitations of earlier methods and pointed the way towards a more flexible estimation strategy. In addition, new research designs have been proposed for identifying causal mechanisms.

In this paper, we introduce a full featured R package, **mediation**, for studying causal mechanisms. The **mediation** package allows users to (1) investigate the role of causal mechanisms using different types of data and statistical models, (2) explore how results change as identification assumptions are relaxed, and (3) calculate quantities of interest under alternative research designs. We focus on the demonstration of the functionalities available through the **mediation** package. The statistical theory that underlies the procedures implemented in the

mediation package is presented elsewhere along with various empirical examples (Imai, Keele, and Yamamoto 2010c; Imai, Keele, Tingley, and Yamamoto 2011; Imai, Keele, and Tingley 2010a; Imai, Tingley, and Yamamoto 2012).

The **mediation** package is freely available for download via the Comprehensive R Archive Network (CRAN) and runs on a variety of computing platforms. In addition, a Stata version of the package is available but has a more limited functionality (Hicks and Tingley 2011). The first version of the **mediation** package appeared at CRAN in 2009, and Imai, Keele, Tingley, and Yamamoto (2010b) discuss an earlier version of the package. Over the last three years, however, we have dramatically improved the package, which as a result has a significant number of new functionalities and improvements. The current paper thus provides an up-to-date description of the analyses that can be conducted via the **mediation** package. To install the **mediation** package, use the following standard syntax for installing an R package,

```
> install.packages("mediation")
```

where users may be prompted with the selection of CRAN mirror from which the package will be downloaded. This step needs to be done only once (unless one wishes to update the **mediation** package to the new version).

In the next section, we present an overview of the **mediation** package. We then describe the functionalities of the package for the model-based causal mediation analysis (Section 3), the design-based causal mediation analysis (Section 4), and the analysis of causally dependent multiple mediators (Section 5). Finally, Section 6 concludes.

2. Overview of the mediation Package

The **mediation** package consists of several main functions as well as various methods for summarizing output from these functions (e.g. `plot` and `summary`). The package requires little programming knowledge on the user's side. Figure 1 illustrates the basic structure of the **mediation** package, which distinguishes between model-based and design-based inference. Model-based inference has been standard practice in the mediation analysis to date. In the experimental setting, the treatment variable is randomized and the mediating and outcome variables are observed. Imai et al. (2010a) show that a range of parametric and semi-parametric models may then be used to estimate the average causal mediation effect, defined below, and other quantities of interest. This modeling approach relies on the sequential ignorability assumption for point identification, which as Imai et al. (2010a) show, provides a general purpose algorithm for calculation of quantities of interest. In contrast, design-based inference primarily employs the features of the experimental design and does not require the sequential ignorability assumption. The formal identification properties of these designs are studied by Imai et al. (2012) and the examples from experimental and observational studies are contained in Imai et al. (2011, 2012). We refer readers to these papers for the details about the statistical methods implemented via the **mediation** package.

Before describing the functions available in **mediation**, we briefly define the quantities of interest that our software is designed to estimate. Here, we use the potential outcomes framework to define these quantities. Let $M_i(t)$ denote the potential value of a mediator of interest for unit i under the treatment status $T_i = t$. Let $Y_i(t, m)$ denote the potential outcome that would result if the treatment and mediating variables equal t and m, respectively. Consider a

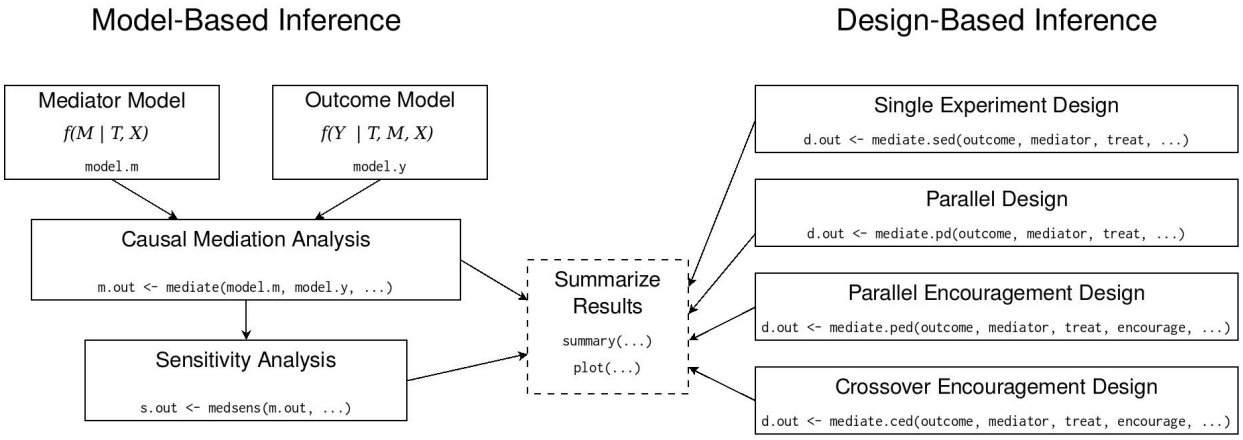

Figure 1: Structure of the **mediation** package as of version 4.0.

standard experimental design where only the treatment variable is randomized. We observe only one of the potential outcomes, and the observed outcome, Y_i, equals $Y_i(T_i, M_i(T_i))$ where $M_i(T_i)$ represents the observed value of the mediator M_i. With this notation, the total unit treatment effect can be written as,

$$\tau_i \equiv Y_i(1, M_i(1)) - Y_i(0, M_i(0)). \tag{1}$$

We can decompose this total effect into the two components. First, the *causal mediation effects* are represented by (Robins and Greenland 1992; Pearl 2001),

$$\delta_i(t) \equiv Y_i(t, M_i(1)) - Y_i(t, M_i(0)), \tag{2}$$

for each treatment status $t = 0, 1$. All other causal mechanisms can be represented by the *direct effects* of the treatment as,

$$\zeta_i(t) \equiv Y_i(1, M_i(t)) - Y_i(0, M_i(t)), \tag{3}$$

for each unit i and each treatment status $t = 0, 1$. Together, we see that they sum up to the total effect,

$$\tau_i = \delta_i(t) + \zeta_i(1 - t) \tag{4}$$

for $t = 0, 1$. The case of multiple candidate mediating variables requires additional notation and is discussed in Section 5. The *average causal mediation effects* (ACME) $\bar{\delta}(t)$ and the average direct effects (ADE) $\bar{\zeta}(t)$, represent the population averages of these causal mediation and direct effects.

Identification of the ACME requires an additional assumption beyond the strong ignorability of the treatment, which is sufficient for identifying the average total effect of the treatment. Let X_i be a vector of the observed pre-treatment confounders for unit i. The key identifying assumption is called sequential ignorability and can be written as,

Assumption 1 (Sequential Ignorability (Imai et al. 2010c))

$$\{Y_i(t', m), M_i(t)\} \perp\!\!\!\perp T_i \mid X_i = x, \tag{5}$$

$$Y_i(t', m) \perp\!\!\!\perp M_i(t) \mid T_i = t, X_i = x, \tag{6}$$

where $0 < \Pr(T_i = t \mid X_i = x)$ *and* $0 < p(M_i = m \mid T_i = t, X_i = x)$ *for* $t = 0, 1$*, and all* x *and* m *in the support of* X_i *and* M_i*, respectively.*

Equation (5) is the standard strong ignorability of the treatment assignment and is satisfied, for example, if the treatment is randomized (possibly conditional on X_i). However, equation (6) requires that the mediator is also ignorable given the observed treatment and pre-treatment confounders. This additional assumption is quite strong because it excludes the existence of (measured or unmeasured) post-treatment confounders as well as that of unmeasured pre-treatment confounders. This assumption, therefore, eliminates the need to consider multiple mediators that are causally related to each other (see Section 5 for the method that is designed to deal with such a scenario).

3. Model-based Causal Mediation Analysis

In this section, we briefly discuss the functionalities of the **mediation** package for the model-based causal mediation analysis under the assumption of sequential ignorability. Many of these functionalities are described in detail in Imai et al. (2010b), but the current version of the package accommodates a larger class of statistical models.

The model-based causal mediation analysis proceeds in two steps. First, the researchers specify two statistical models, the mediator model for the conditional distribution of the mediator M_i given the treatment T_i and a set of the observed pre-treatment covariates X_i and the outcome model for the conditional distribution of the outcome Y_i given T_i, M_i, and X_i. These models are fitted separately and then their fitted objects comprise the main inputs to the `mediate` function, which computes the estimated ACME and other quantities of interest under these models and the sequential ignorability assumption. Since the sequential ignorability assumption is untestable, we recommend that the researchers conduct a sensitivity analysis via the `medsens` function, which can be applied to certain statistical models. We now illustrate these functionalities with an empirical example.

3.1. Estimation of the Average Causal Mediation Effects

The `mediate` function takes various standard model objects (such as `lm` and `glm`), which correspond to mediator and outcome models, and returns the estimates of the average causal mediation effects along with other causal quantities of interest. The output of the `mediate` function can be passed to the `plot` and `summary` functions in order to obtain graphical and numerical summaries, respectively. The `mediate` function automatically detects the type of models used for the mediator and outcome models and calculate the estimates of the ACME and other quantities of interest via the general algorithms described in Imai et al. (2010a). Our estimation strategy overcomes the limitation of the standard methods based on the product or difference of coefficients, which is only applicable when both the mediator and outcome models are linear regressions. In contrast, the algorithms used in the **mediation** package nest

	Outcome Model Types						
Mediator Model Types	Linear	GLM	Ordered	Censored	Quantile	GAM	Survival
Linear (`lm`)	✓	✓	✓*	✓	✓	✓*	✓
GLM (`glm`/`bayesglm`)	✓	✓	✓*	✓	✓	✓*	✓
Ordered (`polr`/`bayespolr`)	✓	✓	✓*	✓	✓	✓*	✓
Censored (tobit via `vglm`)	-	-	-	-	-	-	-
Quantile (`rq`)	✓*	✓*	✓*	✓*	✓*	✓*	✓
GAM (`gam`)	✓*	✓*	✓*	✓*	✓*	✓*	✓*
Survival (`survreg`)	✓	✓	✓*	✓	✓	✓*	✓

Table 1: Types of Statistical Models That Can be Used with the `mediate` Function. Asterisks, *, indicate the model combinations that can only be estimated using the nonparametric bootstrap (i.e. with the argument `boot = TRUE` for the `mediate` function).

this as a special case and accommodates a greater range of statistical models as shown in Table 1.

We now illustrate the use of the `mediate` function with an empirical example based on the `framing` data of Brader, Valentino, and Suhat (2008). This data set is a part of the **mediation** library and can be loaded via the following syntax,

```
> ## load the package
> library(mediation)
> ## load the framing data
> data(framing)
```

A brief description of each variable in the data can be obtained through a help file,

```
> ?framing
```

Brader et al. (2008) conducted a randomized experiment where subjects are exposed to different media stories about immigration and the authors investigated how their framing influences attitudes and political behavior regarding immigration policy. They posit the role of anxiety as the mediating variable for the causal effect of framing on public opinion. We first fit the mediator model where the measure of anxiety (`emo`) is modeled as a function of the framing treatment (`treat`) and pretreatment covariates (`age`, `educ`, `gender`, and `income`). Next, we model the outcome variable, which is a binary variable indicating whether or not the participant agreed to send a letter about immigration policy to his or her member of Congress (`cong_mesg`). The explanatory variables of the outcome model include the mediator, treatment status, and the same set of pre-treatment variables as those used in the mediator model. In this example, the treatment is expected to increase the level of respondents' emotional response, which in turn is hypothesized to make subjects more likely to send a letter to his or her member of Congress. We use the linear least squares regression and the probit regression for the mediator and outcome models, respectively.

```
> ## Mediator Model
> med.fit <- lm(emo ~ treat + age + educ + gender + income, data = framing)
```

```
> ## Outcome Model
> out.fit <- glm(cong_mesg ~ emo + treat + age + educ + gender + income,
+                data = framing, family = binomial("probit"))
```

We now use the `mediate` function to estimate the ACME and average direct effects. As the inputs to this function, we must specify the model fits (in this case `med.fit` and `out.fit`) as well as the names of the treatment and mediating variables, which are represented as the arguments `treat` and `mediator`, respectively. Here, we use the default number of simulations, which is `sims = 1000`, to calculate the uncertainty estimates but one may wish to increase this number if the estimates vary too much from one simulation to another. In addition, we use a robust heteroskedasticity consistent variance-covariance matrix from the **sandwich** package (`vcovHC`). Finally, like most functions in R, the results of the `mediate` function can be summarized numerically by the `summary` function, which calculates point estimates, confidence intervals, and the p-values, for the average direct, indirect, and total effects. The syntax is now given as,

```
> library(sandwich)
> med.out <- mediate(med.fit, out.fit, treat = "treat", mediator = "emo",
+                    robustSE = TRUE)
> summary(med.out)

Causal Mediation Analysis

Quasi-Bayesian Confidence Intervals

                               Estimate 95% CI Lower 95% CI Upper p-value
Mediation Effect_0               0.0822       0.0329       0.1435    0.01
Mediation Effect_1               0.0826       0.0318       0.1435    0.01
Direct Effect_0                  0.0130      -0.0909       0.1348    0.78
Direct Effect_1                  0.0134      -0.0985       0.1447    0.79
Total Effect                     0.0956      -0.0269       0.2281    0.16
Proportion via Mediation_0       0.7917      -3.3546       5.8739    0.37
Proportion via Mediation_1       0.8064      -2.9956       5.3006    0.34
Mediation Effect (Ave.)          0.0824       0.0322       0.1433    0.01
Direct Effect (Ave.)             0.0132      -0.0942       0.1401    0.78
Proportion via Mediation (Ave.)  0.7991      -3.1751       5.5873    0.36

Sample Size Used: 265


Simulations: 1000
```

One new feature in the tabular output from the `mediate` functions is the addition of p-values for the various estimates. In this example, the estimated ACMEs are statistically significantly different from zero but the estimated average direct and total effects are not. The results suggest that the treatment in the framing experiment may have increased emotional response, which in turn made subjects more likely to send a message to his or her member of

Congress. Here, since the outcome is binary all estimated effects are expressed as the increase in probability that the subject sent a message to his or her Congressperson.

In addition, we can use the bootstrap rather than the Quasi-Bayesian Monte Carlo simulation for variance estimation via the `boot = TRUE` argument,

```
> med.out <- mediate(med.fit, out.fit, boot = TRUE, treat = "treat",
+                    mediator = "emo")
> summary(med.out)

Causal Mediation Analysis

Confidence Intervals Based on Nonparametric Bootstrap

                               Estimate 95% CI Lower 95% CI Upper p-value
Mediation Effect_0               0.0850       0.0364       0.1408    0.01
Mediation Effect_1               0.0860       0.0374       0.1404    0.01
Direct Effect_0                  0.0117      -0.1055       0.1327    0.86
Direct Effect_1                  0.0127      -0.1155       0.1407    0.86
Total Effect                     0.0977      -0.0307       0.2274    0.16
Proportion via Mediation_0       0.8699      -6.2027       6.5899    0.30
Proportion via Mediation_1       0.8803      -5.2254       6.0369    0.26
Mediation Effect (Ave.)          0.0855       0.0376       0.1386    0.01
Direct Effect (Ave.)             0.0122      -0.1100       0.1361    0.86
Proportion via Mediation (Ave.)  0.8751      -5.7141       6.3102    0.28

Sample Size Used: 265


Simulations: 1000
```

The output now indicates that the bootstrap is used for inferences. As expected, the results are largely the same. In general, so long as computing power is not an issue, analysts should estimate confidence intervals via the bootstrap with more than 1000 resamples, which is the default number of simulations.

As an alternative to the numerical summary, using the output from the `mediate` function as the input to the `plot` command provides a graphical summary of the three parameters (indirect, direct, and total effects) along with their confidence intervals. Figure 2 shows the result of plotting the `med.out` object.

Finally, it is possible that the ACMEs are conditional on the treatment value. In such a situation, the researcher can add an interaction term between the treatment and mediator to the outcome model. Then, the `mediate` function automatically detects the change in the specification and estimates the ACME conditional on treatment status. In the output given below, the estimated ACME now varies with treatment status.

```
> med.fit <- lm(emo ~ treat + age + educ + gender + income, data=framing)
> out.fit <- glm(cong_mesg ~ emo * treat + age + educ + gender + income,
```

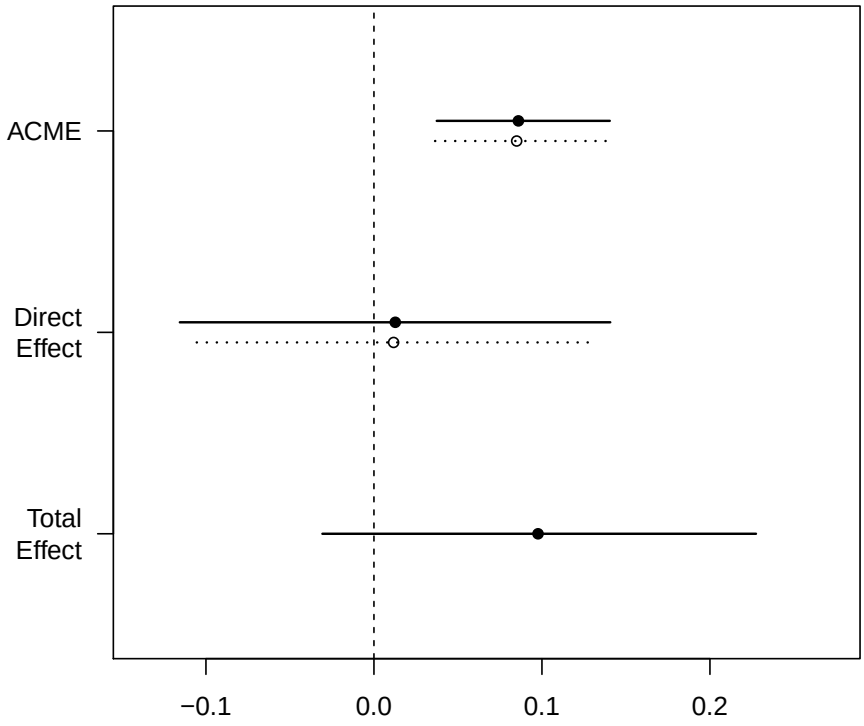

Figure 2: Graphical Display of Results from the `mediate` Function.

```
+                data = framing, family = binomial("probit"))
> med.out <- mediate(med.fit, out.fit, treat = "treat", mediator = "emo",
+                    robustSE = TRUE)
> summary(med.out)

Causal Mediation Analysis

Quasi-Bayesian Confidence Intervals

                              Estimate 95% CI Lower 95% CI Upper p-value
Mediation Effect_0             0.07524      0.02837      0.13009    0.00
Mediation Effect_1             0.09707      0.03765      0.16635    0.01
Direct Effect_0               -0.00424     -0.11980      0.11861    0.98
Direct Effect_1                0.01759     -0.10976      0.15416    0.77
Total Effect                   0.09284     -0.03676      0.22397    0.17
Proportion via Mediation_0     0.72918     -4.26498      8.92530    0.39
Proportion via Mediation_1     0.97391     -4.47265      9.26372    0.35
```

```
Mediation Effect (Ave.)          0.08616      0.03686      0.14262    0.00
Direct Effect (Ave.)             0.00668     -0.11167      0.13517    0.89
Proportion via Mediation (Ave.)  0.85154     -4.28084      9.65900    0.38

Sample Size Used: 265


Simulations: 1000
```

3.2. Moderated Mediation

One important feature of the `mediate` function is the analysis of moderated mediation. Often analysts hypothesize that the magnitude of the ACME depends on (or moderated by) a pre-treatment covariate. Such a pre-treatment covariate is called a moderator. In the framing example, the ACME may be much stronger among older respondents than younger ones. In other words, the ACME may be moderated by age. While earlier versions of the **mediation** package did not allow for this moderated mediation analysis, that functionality has now been added. To use this functionality, the researchers must alter both the statistical models as well as the syntax for the `mediate` function. First, the mediator and outcome models should contain the moderator and its interaction terms with respect to the treatment and mediating variables. For example, we may modify the previous models as follows,

```
> med.fit <- lm(emo ~ treat * age + educ + gender + income, data=framing)
> out.fit <- glm(cong_mesg ~ emo + treat * age + emo * age + educ + gender
+                + income, data = framing, family = binomial("probit"))
```

Once the two models are estimated, the researcher must specify the levels of the moderator at which effects will be calculated by the `mediate` function. In the current example, this can be done by setting the `age` covariate to a specific value. To allow the mediation effects to be moderated by age, we set the value of `age` to be 20 in one model and 60 in another model. More complicated moderated mediations involving multiple moderators could be specified by expanding the list of the covariates.

```
> med.age20 <- mediate(med.fit, out.fit, treat = "treat",
+                      mediator = "emo", covariates = list(age = 20))
> med.age60 <- mediate(med.fit, out.fit, treat = "treat",
+                      mediator = "emo", covariates = list(age = 60))
> summary(med.age20)

Causal Mediation Analysis

Quasi-Bayesian Confidence Intervals

(Inference Conditional on the Covariate Values Specified in `covariates')

                               Estimate 95% CI Lower 95% CI Upper p-value
Mediation Effect_0              0.07224      0.00224      0.16802    0.14
```

```
Mediation Effect_1               0.08775      0.00257      0.18787    0.11
Direct Effect_0                  0.22691      0.00500      0.45770    0.06
Direct Effect_1                  0.24243      0.00553      0.47505    0.04
Total Effect                     0.31466      0.08651      0.54106    0.01
Proportion via Mediation_0       0.21003      0.00662      0.90202    0.34
Proportion via Mediation_1       0.27078      0.01025      0.92015    0.27
Mediation Effect (Ave.)          0.07999      0.00241      0.17873    0.11
Direct Effect (Ave.)             0.23467      0.00526      0.46165    0.05
Proportion via Mediation (Ave.)  0.24040      0.00843      0.91108    0.30

Sample Size Used: 265


Simulations: 1000

> summary(med.age60)

Causal Mediation Analysis

Quasi-Bayesian Confidence Intervals

(Inference Conditional on the Covariate Values Specified in `covariates')

                                Estimate 95% CI Lower 95% CI Upper p-value
Mediation Effect_0                0.0817       0.0170       0.1528    0.02
Mediation Effect_1                0.0734       0.0141       0.1367    0.02
Direct Effect_0                  -0.0861      -0.2166       0.0677    0.20
Direct Effect_1                  -0.0944      -0.2386       0.0697    0.22
Total Effect                     -0.0127      -0.1604       0.1475    0.91
Proportion via Mediation_0       -0.3303     -13.5472      14.7630    0.81
Proportion via Mediation_1       -0.2546     -12.6314      14.0618    0.82
Mediation Effect (Ave.)           0.0776       0.0164       0.1449    0.02
Direct Effect (Ave.)             -0.0902      -0.2256       0.0692    0.21
Proportion via Mediation (Ave.)  -0.2925     -13.1377      14.4166    0.82

Sample Size Used: 265


Simulations: 1000
```

Thus, the researcher receives two different sets of output. In the first output, the average mediation effect is estimated for those who are 20 years old. In contrast, the second output applies to those who are 60 years old.

3.3. Non-Binary Treatment Variables

Experimental manipulations are often complex and go beyond simple treatment and control conditions. In the framing experiment, for example, the researchers actually used a 2×2

	Outcome Model Types	
Mediator Model Types	Linear	Binary Probit
Linear	✓	✓
Binary Probit	✓	-

Table 2: The Types of Models That Can Be Handled by `medsens` for Sensitivity Analysis.

factorial design. That is, each subject was exposed to two different binary treatments, yielding four different experimental manipulations. In the analysis presented above, we have focused on a comparison of one of these conditions relative to the other three conditions. The `mediate` function, however, has the capability to handle more complex experimental contrasts, which can be represented by a non-binary treatment variable.

Here, instead of using the binary `treat` variable, we use a variable named `cond`, which records which of the four conditions the subject was randomly exposed to. Using the `control.value` and `treat.value` options, the user can calculate the specific contrast of interest. For example, the comparison between the second and third conditions can be done with the following code.

```
> med.fit <- lm(emo ~ cond + age + educ + gender + income, data = framing)
> out.fit <- glm(cong_mesg ~ emo + cond + age + educ + gender + income,
+                data = framing, family = binomial("probit"))
> med23.out <- mediate(med.fit, out.fit, treat = "cond", mediator = "emo",
+                      control.value = 2, treat.value = 3)
> summary(med23.out)
```

Similarly, the researcher can compare the first and fourth experimental conditions via the following syntax,

```
> med14.out <- mediate(med.fit, out.fit, treat = "cond", mediator = "emo",
+                      control.value = 1, treat.value = 4)
> summary(med14.out)
```

Nothing changes in the format of the output, but the contrasts differ depending on the categories chosen for comparison by the researcher. In the case of a continuous treatment variable, the researcher would specify two values of the treatment to make the contrast (Imai et al. 2010a). For example, the causal mediation effects can be defined for any two levels of the treatment,

$$\delta_i(t; t_1, t_0) \quad \equiv \quad Y_i(t, M_i(t_1)) - Y_i(t, M_i(t_0)), \tag{7}$$

where $t_1 \neq t_0$. The corresponding average causal mediation effect is defined as $\bar{\delta}(t; t_1, t_0) \equiv \mathbb{E}(\delta_i(t; t_1, t_0))$. Thus, the researcher can set `control.value` to t_0 and `treat.value` to t_1. The researcher may also vary the value of t_1, while fixing the base line value of t_0, to examine how the ACME changes as the function of t_1.

3.4. Sensitivity Analysis for Sequential Ignorability

Sequential ignorability is a strong assumption, and therefore a sensitivity analysis is recommended. The **mediation** package allows the researcher to conduct a sensitivity analysis for

the possible existence of unobserved pre-treatment covariates. Specifically, the output of the `mediate` function can be passed to the `medsens` function, which then computes the values of causal quantities as a function of sensitivity parameters. Both `summary` and `plot` functions are available for sensitivity analysis, and they display the results in the tabular and graphical form, respectively. Since derivation of sensitivity formulas must be done on a case-by-case basis, the range of options for conducting sensitivity analyses is somewhat limited. Table 2 gives the model combinations currently supported by the `medsens` function.

In our running example, after computing the ACME, we conduct a sensitivity analysis by passing the object from `mediate` to the `medsens` function. We first choose as the sensitivity parameter the correlation ρ between the residuals of the mediator and outcome regressions (Imai et al. 2010c,a). If there exist unobserved pre-treatment confounders which affect both the mediator and the outcome, we expect that the sequential ignorability assumption is violated and ρ is no longer zero. The sensitivity analysis is conducted by varying the value of ρ and examining how the estimated ACME changes. The following syntax can be used to conduct this analysis,

```
> ## Mediation model
> med.fit <- lm(emo ~ treat + age + educ + gender + income, data = framing)
> ## Outcome model
> out.fit <- glm(cong_mesg ~ emo + treat + age + educ + gender + income,
+                data = framing, family = binomial("probit"))
> ## Computing the ACME etc.
> med.out <- mediate(med.fit, out.fit, treat = "treat", mediator = "emo",
+                    robustSE = TRUE)
> ## Sensitivity analysis
> sens.out <- medsens(med.out, rho.by = 0.1, effect.type = "indirect")
> ## Summary output
> summary(sens.out)

Mediation Sensitivity Analysis: Average Mediation Effect

Sensitivity Region: ACME for Control Group

     Rho Med. Eff. 95% CI Lower 95% CI Upper R^2_M*R^2_Y* R^2_M~R^2_Y~
[1,] 0.3    0.0056      -0.0075        0.016         0.09       0.0493
[2,] 0.4   -0.0090      -0.0276        0.002         0.16       0.0877

Rho at which ACME for Control Group = 0: 0.3
R^2_M*R^2_Y* at which ACME for Control Group = 0: 0.09
R^2_M~R^2_Y~ at which ACME for Control Group = 0: 0.0493


Sensitivity Region: ACME for Treatment Group

     Rho Med. Eff. 95% CI Lower 95% CI Upper R^2_M*R^2_Y* R^2_M~R^2_Y~
[1,] 0.3    0.0064      -0.0086       0.0196         0.09       0.0493
[2,] 0.4   -0.0109      -0.0334       0.0019         0.16       0.0877
```

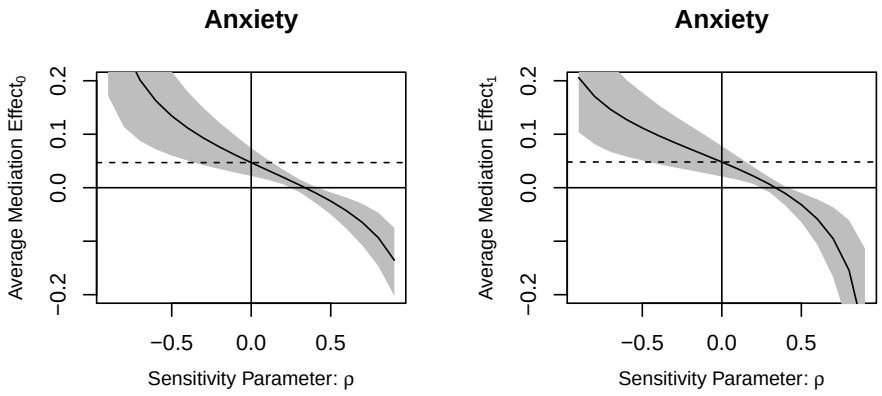

Figure 3: Graphical Display of Results from the `medsens` Function. Results as a Function of ρ.

```
Rho at which ACME for Treatment Group = 0: 0.3
R^2_M*R^2_Y* at which ACME for Treatment Group = 0: 0.09
R^2_M~R^2_Y~ at which ACME for Treatment Group = 0: 0.0493
```

where `rho.by = 0.1` specifies that ρ will vary from -0.9 to 0.9 by 0.1 increments, and `effect.type = "indirect"` means that sensitivity analysis is conducted for the ACME. Alternatively, specifying `effect.type = "direct"` performs sensitivity analysis for the ADE and `"both"` returns sensitivity analysis for the ACME and ADE.

The tabular output from the `summary` function displays the values of ρ at which the confidence intervals contain zero for the ACME. For both the control and treatment conditions, the confidence intervals for the ACME contain zero when ρ equals 0.3 and 0.4. An alternative but mathematically equivalent way to conduct sensitivity is in terms of the product of R^2 (or coefficients of determination) statistics from the mediator and outcome models. Discussed in more detail elsewhere (Imai et al. 2010c, 2011, 2010a), the first row captures the point at which the ACME is 0 as a function of the proportions of residual variance in the mediator and outcome explained by the hypothesized unobserved confounder. The second line uses the total variance instead of residual variance. We use R^{*2} for residual variance and $\tilde{R}^2$ for total variance. For example, when the product of the original variance explained by the omitted confounding is .049 the point estimate for ACME would be 0.

A graphical display is often more intuitive and useful for the sensitivity analysis, especially for the R^2 interpretations. This can be done, as before, by passing the object from the `medsens` function to the `plot` function. The `plot` function allows the researcher to graphically summarize the results of sensitivity analysis either in terms of ρ (`sens.par = "rho"`) or R^2 statistics (`sens.par = "R2"`).

```
> plot(sens.out, sens.par = "rho", main = "Anxiety", ylim = c(-0.2, 0.2))
```

When using the R^2 statistic version of sensitivity analysis the user must specify whether

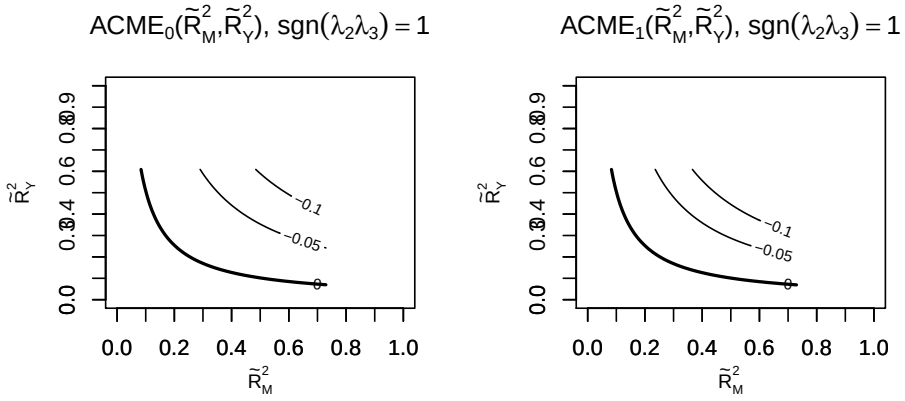

Figure 4: Graphical Display of Results from the `medsens` Function. Results as a Function of $\tilde{R}^2$.

the hypothesized confounder affects the the mediator and outcome variables in the same direction or in different directions. This matters because the sensitivity analysis is in terms of the product of R^2 statistics. In the current example, we assume that the confounder influences both variables in the same direction by setting `sign.prod = "positive"` (rather than `sign.prod = "negative"`). Here, we plot the total variance version of the sensitivity analysis. The bold line represents the various combinations of the R^2 statistics where the ACME would be 0 (in this case the product equals .049). The graphical display also presents the corresponding contour plots for other products of the R^2 statistics.

```
> plot(sens.out, sens.par = "R2", r.type = "total", sign.prod = "positive")
```

4. Design-Based Causal Mediation Analysis

An alternative approach to model-based inference is to use different research designs that are specifically designed for identifying causal mechanisms. Imai et al. (2012) propose several such designs and describe the assumptions required for the identification of causal mediation effects under each of the designs. In this section we briefly illustrate how to use our software to calculate the estimates of the quantities of interest under each design.

4.1. Single Experiment Design

The single experiment design randomizes the treatment variable and measures the mediating and outcome variables. In Section 3, we discussed estimation functions that can be used with parametric and semi-parametric models. If the researchers wish to pursue a completely non-parametric approach the **mediation** package offers two options via the `mediate.sed` function. First, the researchers can continue to make the sequential ignorability assumption and non-parametrically estimate the ACME. This approach works only when the mediator variable is discrete. Second, the sharp bounds on the ACME can be computed under the assumption

that only the treatment is randomized. Unfortunately, Imai et al. (2012) derive the bounds in the case with all binary variables (treatment, mediator, and outcome) and show that the bounds are never informative about the sign of the ACME (i.e., they always cover 0).

Most mediation analysis proceeds under the sequential ignobility assumption. Those analyses also tend to be model-based, but they need not be. Imai et al. (2010c) outline a design-based estimator for the ACME for when the mediator is discrete. This estimator for the ACME is fully nonparametric. One drawback to this estimator is that one can encounter mediator-treatment combinations for which there are no subjects because of data sparsity. Standard error calculation for this estimator is based on either the Delta method or the nonparametric bootstrap.

The `mediate.sed` function requires the names of the outcome, mediator, and treatment variables, along with the name of the data frame that contains these variables. When `SI = TRUE`, the function will return the point estimates under the sequential ignorability assumption, and otherwise the results will be a set of sharp bounds for the ACME. The method for inference also differs slightly from the `mediate` function. When `boot = TRUE` the bootstrap is used, but when `boot = FALSE`, the Delta method is used to compute standard errors.

Below, we present an example using the framing data from Brader et al. (2008). The treatment variable is the same as before, i.e., `treat`, and the mediator is `anx`, which refers to a subject's reported level of anxiety. This four level measure is one component of the `emo` variable that was previously used as the mediator and in the data all treatment-mediator combinations are present (a requirement for the estimator). The outcome variable in this example is `english` and measures on a four point scale how much someone supports English only laws, from strongly support to strongly oppose. Note that the `mediate.sed` function only take numeric variables as arguments. Variables that are stored as factors must be converted to numeric variables as we show below.

```
> framing$english <- as.numeric(framing$english)
> framing$anx <- as.numeric(framing$anx)
> sed.est <- mediate.sed("english", "anx", "treat", data = framing, SI = TRUE,
+                        boot = TRUE)
> summary(sed.est)


Design-Based Causal Mediation Analysis

Single Experiment Design with Sequential Ignorability

Confidence Intervals Based on Nonparametric Bootstrap

Mediation Effect_0:  0.1021 95 % CI  -0.5710  0.7965
Mediation Effect_1:  0.07066 95 % CI  -0.2044  0.3258
Sample Size Used:    265
```

The results from the `summary` function display the mediation effects along with the default 95% confidence intervals. In this example both $\bar{\delta}(0)$ and $\bar{\delta}(1)$ are not significantly different from 0.

4.2. Parallel Design

An alternative to the single experiment design is the "parallel design" proposed by Imai et al. (2012). In this design there are two separate experiments that are run in parallel with subjects randomly assigned to one of the two experiments. The first experiment follows the single experiment design. In the second experiment, subjects are randomly assigned to treatment or control. Then, a randomly selected set of subjects in each condition is assigned a value of the mediating variable. A key assumption of this design is that the manipulation of the mediating variable is possible and has no direct effect on the outcome variable.

Under the parallel design, the ACME is not point identified without an additional assumption. The **mediation** package offers two options via the `mediate.pd` function. First, the researchers can assume no interaction between the treatment and mediating variables by setting `NINT = TRUE`. In this case, the `mediate.pd` function will calculate the ACME along with its bootstrap confidence intervals. Second, the assumption of no-interaction between treatment and mediator can be dropped via `NINT = FALSE`, and then the sharp bounds can be calculated for the ACME. These bounds may be informative about the sign (i.e., do not cover 0) and are always narrower compared to the bounds under the single experiment design where the only assumption is randomization of the treatment.

For illustration, we simulated data based on the media framing experiment by Brader et al. (2008) by creating a population distribution of potential mediators and outcomes (see Imai et al. (2012) for more details). We then sampled 1000 cases from this distribution. In this example, `out` represents the outcome variable (immigration attitudes), `med` represents the mediator (anxiety), and `ttt` represents the treatment. All variables are binary. `manip` represents whether the subject had the mediator manipulated (-1 if mediator is manipulated down, 0 if no manipulation, and 1 if manipulated up). First, the no-interaction assumption is made and options for the number of bootstrap simulations and confidence intervals are specified. In this case, the mediation effect is estimated at -0.12 with 95% confidence intervals spanning $[-0.21, -0.03]$. In the second example, the no interaction assumption is dropped and the sharp bounds are calculated to span $[-0.3, 0.3]$ for the control condition and $[0.2, 0.77]$ for the treatment condition.

```
> data(boundsdata)
> pd <- mediate.pd("out", "med", "ttt", "manip", boundsdata,
+                  NINT = TRUE, sims = 1000, conf.level = 0.95)
> summary(pd)

Design-Based Causal Mediation Analysis

Parallel Design (with No Interaction Assumption)

Mediation Effect:  -0.1236 95 % CI  -0.21944 -0.03273
Sample Size Used:  1000

> pd1 <- mediate.pd("out", "med", "ttt", "manip", boundsdata, NINT = FALSE)
> summary(pd1)

Design-Based Causal Mediation Analysis
```

```
Parallel Design (Interaction Allowed)

                      Lower   Upper
Mediation Effect_0:  -0.3207  0.3303
Mediation Effect_1:  0.2006 0.7680
Sample Size Used:    1000
```

4.3. Parallel Encouragement Design

In many situations, perfect manipulation of the mediating variable may be difficult. In the parallel encouragement design, subjects are split into two separate experiments. The first experiment is based on the single experiment design. In the second experiment subjects are randomly assigned to the treatment and control conditions and then, within each condition, a subset of subjects are randomly encouraged to have a high or low value of the mediator. Both the mediator and outcome variable are then measured. The `mediate.ped` function reports the sharp bounds on two estimands. First is the ACME and second is the ACME for the "compliers" who respond to the encouragement. The calculation of these bounds is accomplished via a standard linear programming approach as discussed in Imai et al. (2012).

The parallel encouragement design requires the analyst to specifically design some form of encouragement. The functionality of the `mediate.ped` closely mirrors that of `mediate.sed`. The key difference is that the analyst must also include an indicator for encouragement. For illustration, we simulated data based on the media framing experiment by Brader et al. (2008). We did this by creating a population distribution of potential mediators and outcomes, and compliance types. We then randomly draw the joint probabilities of the causal types and assign an encouragement status for those in the encouragement condition (see Imai et al. (2012) for more details). Based on the encouragement condition and encouragement status (`enc`, -1 if mediator is encouraged down, 0 if no encouragement, and 1 if encouraged up), the observed binary values of the mediator (`med.enc`) and outcome (`out.enc`) are determined. Using this simulated data we can then pass it through the `mediate.ped` function for the parallel encouragement design.

```
> data(boundsdata)
> ped <- mediate.ped("out.enc", "med.enc", "ttt", "enc", boundsdata)
> summary(ped)


Design-Based Causal Mediation Analysis

Parallel Encouragement Design

                       Lower   Upper
Population ACME_0:  -0.4341  0.3237
Complier ACME_0     -0.1465  0.2077
Population ACME_1:  -0.02014  0.74339
Complier ACME_1:    0.01137 0.70721
Sample Size Used:   1000
```

Here, the results from `mediate.ped` function are a set of sharp bounds. We see that for the compliers, the sharp bounds on ACME under the treatment condition are informative as they do not cross 0.

4.4. Crossover Encouragement Design

The fourth experimental design included in the **mediation** package is the crossover encouragement design. Under this design, subjects are exposed to two experiments. In the first experiment, the treatment variable is randomized and the mediator and outcome variables observed. In the second experiment, the treatment condition is set to the opposite value from the first period, but an encouragement is given to a randomly selected set of subjects so that the mediator variable will take on the value observed in the first experiment. Under this design, the ACME is point identified for the set of subjects that are able to have their mediator value manipulated (known as "pliable units"). A crucial identification assumption is that the first experiment does not influence behavior in the second experiment. For this experimental design the `mediate.ced` function calculates point estimates and the bootstrap is used for estimates of uncertainty.

For illustration, we simulated data based on the identification assumptions necessary for this design. `Y2` is the value of the outcome variable in the second experiment, `M1` and `M2` are the mediator values for the first and second experiment, `T1` is the value of the treatment in the first experiment, and `Z` indicates whether the subject's mediator value in the second experiment is encouraged to take on the value opposite to that observed in the first experiment. All variables are binary.

```
> data(CEDdata)
> ced <- mediate.ced("Y2", "M1", "M2", "T1", "Z", CEDdata, sims = 100)
> summary(ced)

Design-Based Causal Mediation Analysis

Crossover Encouragement Design

Pliable ACME_0:    0.09069 95 % CI  -0.1315  0.3134
Pliable ACME_1:    0.1194 95 % CI  -0.0421  0.3194
Sample Size Used:  2000
```

The results from the `mediate.ced` function are point estimates and confidence intervals for the ACME under the treatment and control conditions. These estimates apply only to the pliable units. In this example, both values of the ACME are positive but the 95% confidence intervals overlap with zero.

5. Analysis of Causally Dependent Multiple Mechanisms

Our discussion so far has focused on a single mediator, M. Frequently, however, researchers take measurements of more than one mediating variable in a given study. Accounting for alternative mechanisms is indeed crucial for the identification of the mechanism of primary

interest, especially when such mechanisms are causally not independent. This is because the alternative dependent mediators affect both the mediator of primary interest and the outcome variable, which, by definition, violates the sequential ignorability assumption (Assumption 1).

5.1. The Methodology

Imai and Yamamoto (2011) develop a framework for dealing with multiple, causally related mediators (or equivalently, post-treatment mediator-outcome confounders). We briefly review this framework. First, let $W_i(t)$ denote the vector of the potential values of those alternative mediators given treatment status t. To allow the causal dependence of both the primary mediator and outcome on W, we write the potential mediator and outcome as $M_i(t, w)$ and $Y_i(t, m, w)$, respectively. The observed values of these potential response variables can then be expressed as $W_i = W_i(T_i)$, $M_i = M_i(T_i, W_i(T_i))$, and $Y_i = Y_i(T_i, M_i(T_i, W_i(T_i)), W_i(T_i))$. The causal mediation effects can now be reexpressed using these notations as,

$$\delta_i(t) \quad = \quad Y_i(t, M_i(1, W_i(1)), W_i(t)) - Y_i(t, M_i(0, W_i(0)), W_i(t)),$$

for $t = 0, 1$. Note that this quantity represents the treatment effects that are transmitted through the mediator of primary interest M_i, irrespective of whether they also come through the alternative mediators W_i or not. Therefore, the quantity of interest remains unchanged from the previous sections, except that the existence of the other mediators are now explicitly taken into consideration.

The framework of Imai and Yamamoto (2011) is based on the following varying coefficient linear structural equations model,

$$M_i(t, w) \quad = \quad \alpha_2 + \beta_{2i} t + \xi_{2i}^\top w + \mu_{2i}^\top tw + \lambda_{2i}^\top x + \varepsilon_{2i}, \tag{8}$$

$$Y_i(t, m, w) \quad = \quad \alpha_3 + \beta_{3i} t + \gamma_i m + \kappa_i tm + \xi_{3i}^\top w + \mu_{2i}^\top tw + \lambda_{3i}^\top x_i + \varepsilon_{3i}, \tag{9}$$

where $\mathbb{E}(\varepsilon_{2i}) = \mathbb{E}(\varepsilon_{3i}) = 0$ without loss of generality. Although these equations may resemble a traditional linear structural equations model at a first glance, they are considerably more flexible because the coefficients are all allowed to vary arbitrarily across individual units.

Imai and Yamamoto (2011) propose two strategies for the analysis of the average causal mediation effects, $\bar{\delta}(t) \equiv \mathbb{E}(\delta_i(t))$. First, it can be shown that the ACME is point identified under the above model and sequential ignorability (a weaker version allowing for post-treatment confounding; see Imai and Yamamoto) if the *homogeneous interaction* assumption is satisfied. This additional assumption is formally written as,

$$Y_i(1, m, W_i(1)) - Y_i(0, m, W_i(0)) \quad = \quad B_i + Cm$$

for any m. The assumption states that the degree of interaction between the treatment and the primary mediator is constant across individual units, which may or may not be plausible depending on the empirical context.

Second, when this assumption is violated, one can express the sharp bounds on the ACME as functions of a parameter representing the degree of the violation, and conduct a sensitivity analysis. The sensitivity parameter here is the standard deviation of the coefficient on the treatment-mediator interaction term, i.e.,

$$\sigma \quad \equiv \quad \sqrt{\mathbb{V}(\kappa_i)},$$

and the expression for the sharp bounds are given in Imai and Yamamoto (2011, footnote 6). Researchers can then analyze robustness to the potential violation of the homogeneous interaction assumption by examining how the location and width of the bounds vary as σ changes.

The sensitivity analysis can also be formulated in terms of two alternative sensitivity parameters, both based on coefficients of determination as in the single mediator case (see Section 3.4). Specifically, we use the proportion of the residual or total variance of the outcome variable that would be explained by allowing the heterogeneity in the treatment-mediator interaction in the outcome model. These parameters are formally defined as

$$R^{2*} = \frac{\mathbb{V}(\tilde{\kappa}_i T_i M_i)}{\mathbb{V}(\eta_{3i}(T_i, M_i, W_i))} \quad \text{and} \quad \widetilde{R}^2 = \frac{\mathbb{V}(\tilde{\kappa}_i T_i M_i)}{\mathbb{V}(Y_i)}, \tag{10}$$

where $\tilde{\kappa}_i = \kappa_i - \mathbb{E}(\kappa_i)$. Researchers may find these parameters to be easier to interpret in substantive terms, as they represent how important it would be to incorporate the interaction heterogeneity in order to explain the variation in the outcome model. Imai and Yamamoto (2011) show that these parameters have a one-to-one relationship with σ, implying that the ACME can also be written as a function of R^{2*} or $\widetilde{R}^2$.

5.2. Single Experiment Design

The above framework has been implemented in the **mediation** package as the `multimed` function. The function takes a data frame containing the necessary variables (outcome, primary mediator, alternative mediators, treatment, and pre-treatment covariates if any) and outputs an object of class `multimed`, a list consisting of estimated bounds along with uncertainty estimates. Note that the theoretical framework accommodates more than one post-treatment confounder, and the current version of `multimed` accommodates multiple post-treatment confounders (version 4.2 or later).

The functionality of `multimed` differs in important ways from `mediate`. First, there is not a separate function for sensitivity analysis. Instead, a sensitivity analysis is conducted within the function along with the estimates of the mediation effects. Second, the arguments for the `multimed` function are rather different. Here, the names of the outcome (`outcome`), primary mediator (`med.main`), alternative mediators (`med.alt`) and treatment (`treat`) variables are passed to the function along with a vector of the names of the pre-treatment covariates to condition on (`covariates`). In the `multimed` function, inference can only be done with the nonparametric bootstrap.

To illustrate the use of the function we revisit the media framing example in Section 3. Here, we use a different outcome variable `immigr`, which is a five category measure of whether immigration should be increased or decreased (treated as a continuous measure). The main mediator is the same composite measure of anxiety, `emo`, and the treatment and pretreatment covariates are defined as before. We now introduce an alternative mediator `p_harm`, which is an eight category measure of the perceived economic harm of immigrants. The reasoning behind the inclusion of this variable is that the media framing treatment may also affect participants' opinion about immigrants by changing their factual belief about the economic impact of increased immigration, which may also affect the level of anxiety and therefore confound the mediator-outcome relationship.

```
> Xnames <- c("age", "educ", "gender", "income")
```

```
> m.med <- multimed(outcome = "immigr", med.main = "emo", med.alt = "p_harm",
+                   treat = "treat", covariates = Xnames,
+                   data = framing, sims = 1000)
> summary(m.med)

Causal Mediation Analysis with Confounding by an Alternative Mechanism

Estimates under the Homogeneous Interaction Assumption:
              Estimate CI lower CI upper
ACME(treated)   0.0645  -0.0851     0.21
ACME(control)   0.1240   0.0221     0.23
ACME(average)   0.1087   0.0151     0.20
Total           0.4175   0.1742     0.66

Sensitivity Analysis:
Values of the sensitivity parameters at which ACME first crosses zero:
              sigma(bounds) sigma(CI) R2s(bounds) R2s(CI) R2t(bounds) R2t(CI)
ACME(treated)        0.0299    0.0000      0.0300  0.0000      0.0178    0.00
ACME(control)        0.0489    0.0173      0.0800  0.0100      0.0474    0.01
ACME(average)        0.0423    0.0173      0.0600  0.0100      0.0356    0.01
```

The `summary` function produces two tables. The first table shows the estimated ACME and total treatment effect and their confidence intervals (default at 95%) under the homogeneous interaction assumption. Three variants of the ACME are shown: the ACME conditional on the treatment group, the control group, and the weighted average of the two with the weights being equal to the proportions of the treatment and control groups. The second table shows key summary results from the sensitivity analysis with respect to possible heterogeneity in treatment-mediator interactions. Specifically, the table presents the values of σ (column 1), R^{2*} (column 3), and $\widetilde{R}^2$ (column 5) at which the true estimated ACMEs equal zero. The remaining columns (2, 4 and 6) show those values for the confidence bands of the three ACMEs.

The results from the `multimed` function can also be analyzed graphically using the `plot` function. One can produce two types of plots, corresponding to the two tables in the `summary` output. First, one can plot the point estimates under the homogeneous interaction assumption by setting the `type` argument to `"point"`, as shown below. The output is in Figure 5.

```
> plot(m.med, type = "point")
```

Second, the results from the sensitivity analysis with respect to σ, R^{2*} or $\widetilde{R}^2$ can be plotted. In this case, the `type` argument can be used to specify which parameter(s) the estimated ACME should be plotted against. The possible values are `"sigma"`, `"R2-residual"`, or `"R2-total"`. One can also choose the types of the ACME from `"treated"`, `"control"` and `"average"` via the `tgroup` argument. In the example below, we plot the estimated ACME for both treatment and control conditions as a function of σ and $\widetilde{R}^2$. The output is in Figure 6.

```
> oldpar <- par(mfrow = c(2,2))
> plot(m.med, type = c("sigma", "R2-total"), tgroup = c("treated", "control"))
> par(oldpar)
```

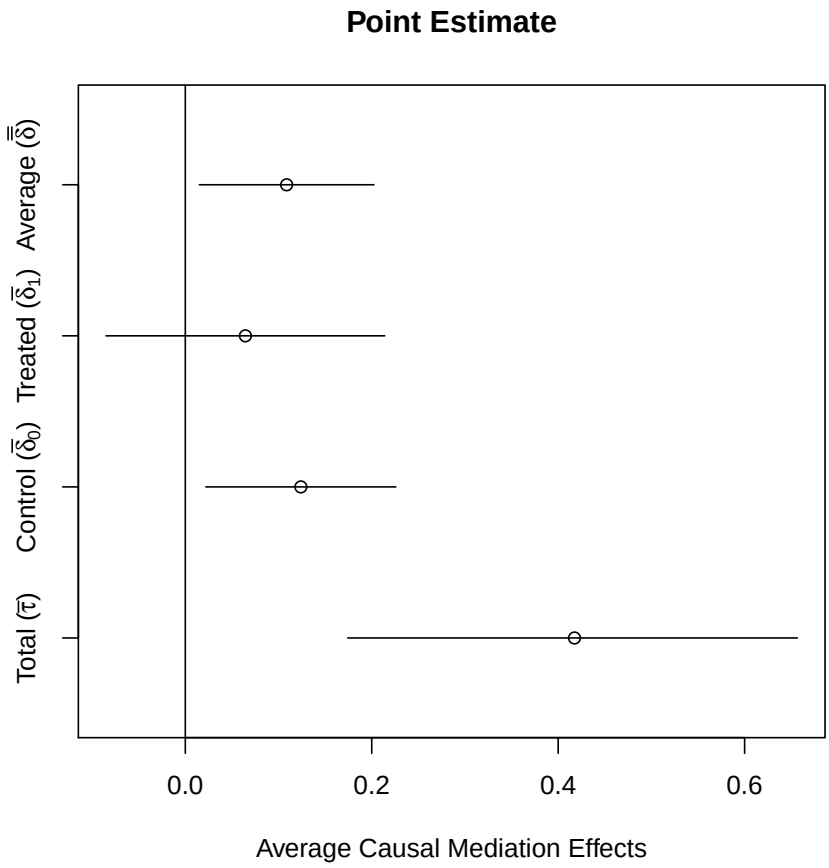

Figure 5: Graphical Summary of the Results from the `multimed` Function under the Homogeneous Interaction Assumption.

5.3. Parallel Design

Imai and Yamamoto (2011, Section 7) show that the above framework can also be applied to the data collected under the parallel design. As discussed in Section 4.2, the parallel design consists of two separate experiments to which subjects are randomly selected. In one experiment, only the treatment is randomized and the researcher observes the mediator and outcome variables, whereas in the other experiment both the treatment and mediator are randomly manipulated and the outcome variable is measured and recorded.

Unlike the single experiment design, one need not assume any kind of sequential ignorability under the parallel design. This is due to the existence of the second experimental group where both the treatment and mediator are randomly assigned. This implies that there is no need to explicitly incorporate an alternative mediator in the analysis, for any kind of post-treatment confounding (observed or unobserved) is allowed to exist in the natural causal process to identify the ACME under the parallel design using the proposed framework.

To apply the framework for the parallel design, one can use the `multimed` function with slightly

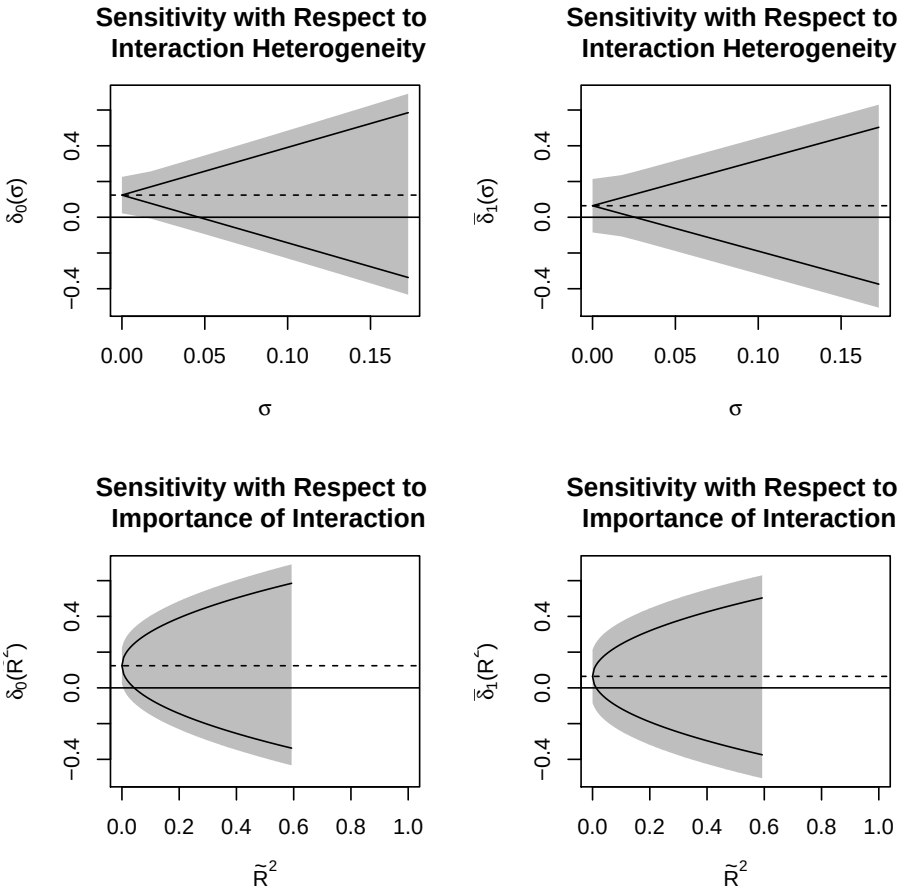

Figure 6: Graphical Summary of Sensitivity Analysis using the `multimed` Function. Results as a Function of σ and $\widetilde{R}^2$.

modified inputs. First, the `med.alt` is no longer needed because the estimation framework is agnostic about what particular alternative mechanisms are confounding the mediator-outcome relationship. Second, one need to supply an additional variable (`experiment`) indicating whether units are assigned to the experiment with (`1`) or without (`0`) mediator manipulations. Finally, the `design` argument must be set to `"parallel"`, as opposed to the default value of `"single"`. For illustration, we again use the simulated data introduced in Section 4.2 and apply the varying coefficient linear structural equations framework.

```
> m.med.para <- multimed(outcome = "out", med.main = "med", treat = "ttt",
+                        experiment = "manip", design = "parallel",
+                        data = boundsdata, sims = 1000)
> summary(m.med.para)

Causal Mediation Analysis with Confounding by an Alternative Mechanism

Estimates under the Homogeneous Interaction Assumption:
```

```
              Estimate CI lower CI upper
ACME(treated)    0.362    0.235     0.49
ACME(control)    0.307    0.185     0.43
ACME(average)    0.322    0.201     0.44
Total            0.206    0.129     0.27

Sensitivity Analysis:
Values of the sensitivity parameters at which ACME first crosses zero:
              sigma(bounds) sigma(CI) R2s(bounds) R2s(CI) R2t(bounds) R2t(CI)
ACME(treated)         0.779     0.543       0.370   0.180       0.344    0.17
ACME(control)         0.627     0.425       0.240   0.110       0.223    0.10
ACME(average)         0.665     0.462       0.270   0.130       0.251    0.12
```

The `plot` function can also be used in the same manner as in the single experiment case. The key differences between the above analysis and Section 4.2 are threefold. First, the point estimates in the first summary table here only rely on the homogeneous interaction assumption, not the stronger assumption of no interaction. Second, however, the estimates here depend on the additional assumption of additivity, which is embodied in the varying coefficient structural equations model in equations (8) and (9). The additivity assumption may not be plausible in some applications and needs to be carefully examined. Finally, the second summary table shows the result of the sensitivity analysis where the homogeneous interaction assumption is gradually relaxed until an arbitrary interaction heterogeneity is allowed. This may be preferred to the nonparametric bounds approach in Section 4.2, which offers less nuanced information about how robust the point estimates are to the violation of the identification assumption.

6. Concluding Remarks

In this paper, we described the functionalities of the **mediation** package, which allows applied researchers to conduct causal mediation analysis in a variety of settings. The package implements a general algorithm for estimating causal mediation effects with a variety of statistical models. Since the causal mediation analysis under the standard research design requires a strong and untestable assumption, we recommend sensitivity analysis which is also implemented via our package. Finally, this strong identification assumption can be relaxed by adopting alternative research designs and we show how to use our package to conduct causal mediation analysis under those new designs. The literature on causal mediation analysis is fast growing and we expect new methods to be developed in the coming years. We hope that the **mediation** package can serve as a platform to which other researchers can add new methods and functionalities.

References

Baron RM, Kenny DA (1986). "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." Journal of Personality and Social Psychology, **51**(6), 1173–1182.

Brader T, Valentino NA, Suhat E (2008). "What Triggers Public Opposition to Immigration? Anxiety, Group Cues, and Immigration." American Journal of Political Science, **52**(4), 959–978.

Haavelmo T (1943). "The Statistical Implications of a System of Simultaneous Equations." Econometrica, **11**, 1–12.

Hicks R, Tingley D (2011). "Causal Mediation Analysis." Stata Journal, **11**, 609–615.

Imai K, Keele L, Tingley D (2010a). "A General Approach to Causal Mediation Analysis." Psychological Methods, **15**(4), 309–334.

Imai K, Keele L, Tingley D, Yamamoto T (2010b). Causal Mediation Analysis Using R, pp. 129–154. Lecture Notes in Statistics. Springer, New York.

Imai K, Keele L, Tingley D, Yamamoto T (2011). "Unpacking the Black Box: Learning about Causal Mechanisms from Experimental and Observational Studies." American Political Science Review, **105**, 765–789. URL http://imai.princeton.edu/research/mediationP.html.

Imai K, Keele L, Yamamoto T (2010c). "Identification, Inference, and Sensitivity Analysis for Causal Mediation Effects." Statistical Science, **25**(1), 51–71.

Imai K, Tingley D, Yamamoto T (2012). "Experimental Designs for Identifying Causal Mechanisms (with discussions)." Journal of the Royal Statistical Society, Series A (Statistics in Society). URL http://imai.princeton.edu/research/Design.html.

Imai K, Yamamoto T (2011). "Identification and Sensitivity Analysis for Multiple Causal Mechanisms: Revisiting Evidence from Framing Experiments." Working paper available at http://imai.princeton.edu/research/medsens.html.

MacKinnon D (2008). Introduction to Statistical Mediation Analysis. Routledge, New York, NY.

Pearl J (2001). "Direct and Indirect Effects." In Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, pp. 411–420. Morgan Kaufmann, San Francisco, CA.

Robins JM, Greenland S (1992). "Identifiability and exchangeability for direct and indirect effects." Epidemiology, **3**(2), 143–155.

Shadish WR, Cook TD, Campbell DT (2001). Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Houghton Mifflin, Boston.

Affiliation:

Dustin Tingley
Department of Government
1737 Cambridge St, Cambridge, MA, USA
E-mail: dtingley@gov.harvard.edu
URL: http://scholar.harvard.edu/dtingley

Teppei Yamamoto
Department of Political Science
77 Massachusetts Avenue, E53-463, Cambridge, MA, USA
E-mail: teppei@mit.edu
URL: http://web.mit.edu/teppei/www

Luke Keele
Department of Political Science
211 Pond Lab
University Park, PA, USA
Email: ljk20@psu.edu
URL: http://www.personal.psu.edu/ljk20

Kosuke Imai
Department of Politics
Princeton University
Princeton, NJ, USA
E-mail: kimai@Princeton.Edu
URL: http://imai.princeton.edu