

Causal Mediation Analysis

Julian Ashwin

August 2022

1 Introduction

Basic set up:

- T_i is treatment status for i
- $M_i(t)$ is the potential value of a mediator for i under treatment $T_i = t$
- $Y_i(t, m)$ is potential outcome from treatment t and mediator m .

Total treatment effect is therefore

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

This total treatment effect is made up of two components, a causal mediation effect:

$$\delta_i(t) \equiv Y_i(t, M_i(1)) - Y_i(t, M_i(0)).$$

All other mechanisms are represented as the direct effects of the treatment

$$\zeta_i(t) \equiv$$

The average causal mediation effect (ACME) is $\bar{\delta}(t)$ and the average direct effect (ADE) is $\bar{\zeta}(t)$.

Identification of the ACME requires an assumption of sequential ignorability.

$$\{Y_i(t', m), M_i(t)\} \perp\!\!\!\perp T_i | X_i = x \quad (1)$$

which is the standard strong ignorability of the treatment assignment (i.e. conditional on x treatment is effectively randomly assigned).

$$Y_i(t', m) \perp\!\!\!\perp M_i(t) | T_i = t, X_i = x \quad (2)$$

which requires that the mediator is also ignorable given treatment and confounders. This is quite strong as it rules out the possibility of multiple mediators that are related to one another. Later on we will show results if there are multiple mediators.

2 Model-based CMA

Two steps:

1. Specify two statistical models: mediator model for the conditional distribution of M given T and X ; and the outcome model for the conditional distribution of Y given T , M and X . These models are estimated separately.
2. Compute the ACME and other quantities of interest from the mediator and outcome models.

Since the sequential ignorability assumption is untestable, we can conduct a sensitivity analysis.

mediate function takes model objects corresponding to the mediator and outcome models and calculates the ACME following the algorithm described in Imai et al (2010a). These algorithms overcome the limitation of the standard methods based on the product or difference of coefficients, which only work if both mediator and outcome models are linear regressions where T and M enter additively.

Treatment and mediator interaction. It could be that the ACME takes different values depending on the baseline treatment status. For this just include and interaction between treatment and mediator in the outcome model. We then get an ACME that varies with the treatment status.

2.1 Moderated mediation

The magnitude of the ACME may depend on a pre-treatment covariate (called the moderator). There are two alternative routes to analyse moderated mediation with the package.

1. Include the moderator and interactions in the outcome and mediator models, interacted with both T and M . Then specify the levels of the moderator at which to calculate the ACME using the covariates argument.
2. Test the significance of the difference between ACME and ADE between two chosen levels of covariate using the test.modmed function.

2.2 Sensitivity analysis for sequential ignorability

For example, choose as the sensitivity parameter the correlation ρ between residuals of the mediator and outcome regressions. If there are pre-treatment confounders which affect both the mediator and the outcome, the sequential ignorability assumption is violated and $\rho \neq 0$. The sensitivity analysis is conducted by varying the value of ρ and examining how the ACME changes.

In the vignette example, the confidence interval for the ACME contains zero

for $\rho = 0.3$ and $\rho = 0.4$.

This sensitivity analysis seems to capture how the results change if we introduce extra confounders or something? Rather than a test of the sequential ignorability function itself.

I skipped the Section on multilevel data for now.

There is a Section on design-based causal mediation analysis, which is a more fully non-parametric approach.

3 Causally dependent multiple mechanisms

Accounting for alternative mechanisms often crucial for the identification of the mechanism of primary interest.

Imai and Yamamoto (2013) develop methods for dealing with multiple mediators.

- When the mediators are not causally related, this is not that difficult. For each mediator, the other mediators are neither pre-treatment nor post-treatment confounders. The effects can thus be consistently estimated by applying the mediate function successively to each one, ignoring the existence of other mediators.
- When multiple mediators are causally related (or equivalently when one mediator acts as a post-treatment confounder for the other) the analysis requires new assumptions.

Let $W_i(t)$ be the vector of potential values of alternative mediators given treatment status t . Allow the causal dependence of the primary mediator $M_i(t, w)$ and the outcome $Y_i(t, m, w)$ on W . The causal mediation effect is then expressed as

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

Noting that this is the effect transmitted through the primary mediator M_i , irrespective of W_i .

In Imai and Yamamoto (2013), analysis is based on varying coefficient linear SEM

$$M_i(t, w) = \alpha_2 + \beta_{2,i}t + \zeta_{2,i}w + \mu_{2,i}tw + \lambda_{2,i}x + \epsilon_{2,i}$$

$$Y_i(t, m, w) = \alpha_3 + \beta_{3,i}t + \gamma_i m + \kappa_i tm + \zeta_{3,i}w + \mu_{3,i}tw + \lambda_{3,i}x + \epsilon_{3,i}$$

where these are super flexible as the coefficients can vary across individual units. There are two strategies for analysing the average causal mediation effect $\delta(t)$.

1. Show that the ACME is point identified under the above model and sequential ignorability if homogeneous interaction is satisfied:

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

This states that the degree of interaction between treatment and primary mediator m is constant across individuals.

2. When this assumption is violated, we can express bounds on the ACME as functions of a parameter representing the degree of violation and conduct a sensitivity analysis wrt the deviation of the coefficient on the treatment mediator interaction term: $\sigma \equiv \sqrt{Var(\kappa_i)}$.

The SEM framework is implemented as the `multimed` function. This takes a data frame containing the necessary variables (outcome, primary mediator, alternative mediator, treatment, and pre-treatment covariates). This function has pretty different arguments:

- name of outcome
- first mediator
- second mediator
- treatment
- pre-treatment covariates to condition on

As an example, use continuous outcome (`immigr`). Main mediator is the composite measure of anxiety (`emo`), the treatment (`treat`) is exposure to media stories, and the pre-treatment covariates are age, education, gender and income. The alternative mediator is `p_harm`, which is a measure of perceived economic harm.

This produces two tables showing the estimated effects and some sensitivity results.

4 Notes on Imai et al (2011)

Traditional approach to causal mediation is through structural equation models. This relies on untestable assumptions and these are often insufficient - conventional exogeneity assumptions are insufficient for identification of causal mechanisms. This paper makes three contributions:

1. Minimum set of assumptions under experimental and observational studies, showing that conventional exogeneity assumptions are insufficient. Then develop a general algorithm for estimating causal mediation effects, correcting common mistakes.
2. Sensitivity analysis to violations of key assumptions.

3. Provide alternative research designs to enable identification of causal mechanisms under less stringent assumptions.

Illustrate with two empirical examples: media priming and observational studies of incumbency advantage.

Decomposing incumbency effects. Incumbency advantage is one of the most studied topics in electoral politics. Many potential mechanisms like "scare-off/quality effect" - deterring high quality challengers, name recognition and resource advantage.

Potential Outcomes refresher. With binary treatment, there are two potential outcomes until one is realised and the other remains unobserved. So $Y_i(t)$ is the potential outcome for unit i under the treatment status t .

In observational studies, treatments are not randomised, so we often adjust for the observed differences in the pretreatment covariates X_i between the treatment and control groups through regression, matching etc... This assumes that there are no omitted variables that affect both treatment and outcome, so $\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp T_i | X_i = x$. For example, controlling for lagged values of the dependent variable.

Under this framework, the ATE is the average difference in outcome means between treatment and control groups.

Defining Causal Mechanisms as Indirect and Direct Effects. In the Cox and Katz case, challenger quality is the mediator (M) through which the incumbency status (T) causally affects the election outcome (Y). As there are other potential mechanisms, an inferential goal is to decompose the causal effect into indirect (the mediator) and direct (all other mechanisms). The total treatment effect is

$$\tau_i \equiv Y_i(1, M_i(1)) - Y_i(0, M_i(0))$$

The indirect/causal mediation effect is

$$\delta_i(t) \equiv Y_i(t, M_i(1)) - Y_i(t, M_i(0))$$

This represents the indirect effects of the treatment on the outcome through the mediating variable. By fixing the treatment and changing only the mediator we eliminate all other causal mechanisms and isolate the hypothesized mechanism. In the Cox and Katz (1996) study, then the indirect effect $\delta_i(1)$ is the difference between the observed vote share $Y_i(1, M_i(1))$ and the counterfactual vote share $Y_i(1, M_i(0))$, i.e.e the vote share if the candidate would receive if he or she faced a challenger whose quality was at the same level as the challenger he or she would have faced if not an incumbent. This thus isolates the portion of the incumbency advantage due to the quality effect.

All other causal mechanisms are then the direct effects of the treatment

$$\zeta_i(t) \equiv Y_i(1, M_i(t)) - Y_i(0, M_i(t))$$

In the incumbency example, $\zeta_i(t)$ represents the difference in vote share of candidate i with and without incumbency status holding the challenger quality at the level that would be realised if the candidate was an incumbent.

Goal is to decompose the ATE into the ACME and the ADE and assess the relative importance of the hypothesized mechanism.

4.1 Nonparametric Identification under Standard Designs

Standard design - treatment assignment is either randomized or assumed to be random given the pre-treatment covariates. Identifying the causal mechanism requires an additional assumption: the Sequential Ignorability assumption.

$$\{Y_i(t', m), M_i(t)\} \perp\!\!\!\perp T_i | X_i = x \quad (3)$$

$$Y_i(t', m) \perp\!\!\!\perp M_i(t) | T_i = t, X_i = x \quad (4)$$

First, given observed pre-treatment confounders, the treatment assignment is assumed to be ignorable (i.e. statistically independent of potential outcomes and potential mediators) - this is the standard no-omitted-variable bias, exogeneity or unconfoundedness assumption.

Second, we assume that the observed mediator is ignorable given the actual treatment status and pre-treatment confounders. In other words, once we have conditioned on a set of covariates, the mediator status is ignorable. This is similar to the standard assumption that treatment assignment is exogenous in observational studies - so even if treatment is truly randomised then this assumption is still reasonably strong.

In the incumbency advantage example, the first part needs to be considered carefully as the treatment is not randomised - we need to assume that incumbency status is random once we adjust for differences in the previous election outcome and partisanship. Furthermore, we also require that the quality of the challenger in the current election is random once we take into account differences in the incumbency status and past election outcomes and partisanship. For both these cases, there may exist unobserved confounders. It is generally impossible to entirely rule out the possibility that there are some unobserved variables which confound relationships, even after conditioning on observable covariates.

However, if we make these assumptions, then both the ACME and the ADE are nonparametrically identified, so can be identified without any additional distributional or functional form assumptions.

Note that even if both treatment and mediator are randomly assigned, sequential ignorability does not necessarily hold. This is because there is a fundamental difference between the causal effect of the mediator itself and the causal mediation effect. The causal mechanism represents how the effect of *treatment* on outcome is transmitted through the mediator. Identifying the effect of the mediator itself is not sufficient to identify the causal mediation effect.

The sequential ignorability assumption requires that the covariates conditioned on only include *pre-treatment* variables (post-treatment confounders require additional strong assumptions for the ACME to be identified, such as no interaction between the treatment and mediator, Robins (2003)). We cannot condition on premediator confounders if they are affected by the treatment.

If the two causal mediators are unrelated, then the ACME for each mediator can be identified as before. In contrast, if there is a causal relationship between one mediator (M) and another mediator (N), then sequential ignorability is not satisfied even if both treatments are exogenous.

For example, if the media cue effect goes through both emotion and cognitive mechanisms but these are entirely unrelated, then sequential ignorability holds. However, if beliefs about economic costs then go on to affect anxiety, then sequential ignorability does not hold.

Inference and sensitivity analysis under standard designs. Approach here is not tied to any specific statistical model - parametric or non-parametric regressions can be used to model the mediator and outcome variables as the identification assumption does not specify a model. There is also a sensitivity analysis to probe the sequential ignorability assumption by quantifying the degree of possible violation.

Standard approach to estimating mediation effects is to use linear equations. The effect of T on Y

$$Y_i = \alpha_1 + \beta_1 T_i + \xi_1 X_i + \epsilon_{i1}$$

The effect of T on M

$$M_i = \alpha_2 + \beta_2 T_i + \xi_2 X_i + \epsilon_{i2}$$

The effect of T and M on Y

$$Y_i = \alpha_3 + \beta_3 T_i + \gamma M_i + \xi_3 X_i + \epsilon_{i3}$$

The standard method is to estimate ACME using the product of $\beta_2 \gamma$, or alternatively the difference $\beta_1 - \beta_3$. Sequential ignorability here implies zero correlation between ϵ_{i2} and ϵ_{i3} .

Estimation Method. As long as sequential ignorability holds, any statistical model can be used to compute the ACME and ADE. The algorithm has two steps:

1. Fit regression models for mediator and outcome. The mediator is modeled as a function of the treatment and any relevant pre-treatment covariate. This gives two sets of predictions for the mediator - with and without the treatment.
2. Use the outcome model to make potential outcome predictions. If we are interested in the ACME under treatment, predict outcome under treatment using value of mediator predicted in treatment condition, then the outcome under the treatment condition but with the mediator prediction from the control condition. The ACME is then the difference between the outcome predictions using the two different values of the mediator. So this corresponds to the average difference in volatility from fixing treatment status but changing volume between levels with and without a FT article.

Uncertainty can then be computed with bootstraps or Monte Carlo approximations.

Sensitivity analysis. Identifying the causal mechanism requires sequential ignorability, which cannot be tested with the observed data. But we can evaluate the robustness of empirical results to a violation of this assumption. This tests whether a slight violation of the assumption of sequential ignorability would lead to substantively different conclusions.

Following Imai, Keele and Tingley (2010) and Imai, Keele and Yamamoto (2010), define ρ as the correlation between ϵ_{i2} , the error in the mediation model, and ϵ_{i3} , the error in the outcome model. If sequential ignorability holds, all relevant pre-treatment confounders have been conditioned on and so ρ is equal to zero. However, if ρ is non zero then some hidden confounder is biasing the ACME estimate. In other words, there is some factor which both affects the outcome and the mediator in some way. This could for example be if there is genuinely new information contained in the news articles. This could both affect volatility and trading volumes that day. Although ρ is unknown, we can calculate the values of ρ for which the ACME is zero/not significant.

This illustrates how even if the exogeneity assumptions hold for both T and M , we can still get a biased estimate of the ACME if the outcome is not independent of the mediator conditional on treatment and covariates.

The interpretation of ρ is perhaps a little obscure, so instead Imai, Keele and Yamamoto (2010) provide an alternative formulation based on how much the omitted variable would alter the R^2 of the mediator and outcome models. For example, if information in the news articles is important in determining volatility or trading volume, then the model excluding this information will have a much smaller R^2 than the model that includes it. The relative change in R^2

Tomorrow - start with reading the Imai paper Read the Peress paper

5 Causal Mediation for media coverage effect

In this case Y_i is the intraday high-low spread, T_i is media coverage and M_i is trading volume. X_i are the pre-treatment confounders including past price movements and implied volatility, fixed effects and other firm or period level information.

The sequential ignorability assumption then that (i) media coverage is random once we adjust for differences in past performance, implied volatility and firm/time specific factors; (ii) trading volume is random once taking into account differences in media coverage, past performance, implied volatility and firm/time specific factors.

For sequential ignorability to imply identification of the ACME, we can only condition on pre-treatment variables. We cannot condition on pre-mediator confounders if they are affected by the treatment. So, for example, we cannot condition Volume on intra-day return if return is also affected by the media coverage.

If there are multiple causal relationships then sequential ignorability is satisfied only if the two mediators are causally unrelated. In our case this would mean that return magnitude and Volume are causally unrelated.

Including the intra day return thus violates the sequential ignorability assumption, but we can include it as a lower bound on the results.

Table 1:

	AGEP	health	wage	wage_zeros	wage_imp	AGEP.1	health.1	wage.1	wage_zeros.1	wage_imp.1
AGEP	1					1				
health	-0.31	1				-0.293	1			
wage	0.084	0.149	1			0.113	0.132	1		
wage_zeros	-0.118	0.213	1	1		-0.09	0.201	1	1	
wage_imp	-0.053	0.215	1	0.938	1	-0.017	0.214	1	0.918	1

On anticipating high volatility. It is to some extent an open question whether the media *causes* volatility or whether it is very good at anticipating it. However, by controlling for implied volatility, we do get an indication that if the media is anticipating volatility it is more adept at doing so than financial markets are.

Table 2:

	AGEP	health	wage	wage_zeros	wage_imp	AGEP.1	health.1	wage.1	wage_zeros.1	wage_imp.1
AGEP	1					1				
health	-0.287	1				-0.27	1			
wage	0.136	0.12	1			0.115	0.117	1		
wage_zeros	-0.088	0.202	1	1		-0.11	0.195	1	1	
wage_imp	0.009	0.21	1	0.911	1	-0.004	0.204	1	0.913	1