Quality & Quantity
https://doi.org/10.1007/s11135-024-01860-7

The derivative-based approach to nonlinear mediation models: insights and applications

Chiara Di Maria[1] · **Claudio Rubino**[1] · **Alessandro Albano**[1,2]

Accepted: 9 February 2024

Abstract
Traditional mediation analysis has been developed in the context of linear models, enabling the estimation of indirect effects through the product of regression coefficients. However, in the presence of nonlinearities, defining and estimating indirect effects becomes more challenging. While nonlinear mediation models are relatively easy to address in the counterfactual-based framework, very few generalizations to nonlinear associational settings have been proposed. One of the most intuitive is the derivative-based approach that, however, seems not to be widely spread among scholars. In this paper, we deepen such an approach to nonlinear mediation models, clarifying and proposing solutions to some issues which have not been addressed by the previous literature. Specifically, we discussed discrete exposures, binary mediators and extensions of this approach to more complex settings like the multilevel one. We also propose to estimate confidence intervals for the indirect effect within a Bayesian framework and compare its performance to that of other approaches in the literature through a simulation study. Finally, a real data application is presented.

Keywords Mediation analysis · Indirect effect · Derivative-based method · Generalised linear models · Bayesian statistics

1 Introduction

In several applied research fields, it is common that the effect of a variable on a response of interest is not entirely direct, but is transmitted by one or more intermediate variables called *mediators*. Mediation analysis is nowadays a widely spread approach to address

✉ Chiara Di Maria
chiara.dimaria@unipa.it

Claudio Rubino
claudio.rubino@unipa.it

Alessandro Albano
alessandro.albano@unipa.it

[1] Department of Economics, Business and Statistics, University of Palermo, Viale delle Scienze, Building 13, 90128 Palermo, Italy

[2] Sustainable Mobility Center (Centro Nazionale per la Mobilità Sostenibile – CNMS), Milan, Italy

Published online: 02 April 2024

such settings. It was developed by social scientists Baron and Kenny (1986), relying on the path-analytic framework developed by Wright (1934) and subsequently further extended to structural equation models (SEM, Bollen 1989).

The SEM framework assumes linearity; that is, the mediator and the outcome models are assumed to be linear (identity link functions and no interactions). This allows us to estimate the indirect effect of the exposure on the outcome, i.e. the part of the effect conveyed by the mediator, as a product of regression coefficients (Baron and Kenny 1986; Bollen 1989; MacKinnon 2008). This approach takes the name of *product method*, and the obtained indirect effect is interpreted as the change in the outcome associated to a one-unit change in the exposure via the mediator. However, in many real-world applications, linearity may fail to hold because of interaction terms in the models or the presence of not normally distributed variables requiring link functions different from the identity. Consider, for example, a mediational setting where the researcher is interested in investigating the relationship between parental support and depression (measured as a binary variable denoting whether a subject is depressed or not) mediated by another binary variable indicating if a subject has ever experienced bullying. Another example could be a study about how stress levels influence the probability of experiencing memory loss through the number of nighttime awakenings (a discrete variable counting the number of times a person wakes up during the night). It is clear that in both cases, the assumption of normality cannot hold, the mediator being either a binary or a discrete count variable, and the outcome always being binary.

Despite the widespread of such variables in applied research, mediation analysis with nonlinear models has primarily been addressed using non-parametric estimands of the indirect effect based on counterfactuals, typical of a causal framework (Pearl 2012a, b; Gaynor et al. 2019; Doretti et al. 2022). This non-parametric formalisation allows for a certain flexibility in the model specification when one fits parametric models for the mediator and the outcome. However, these estimands require the introduction of a different notation and several, sometimes untestable, assumptions for them being identified (i.e. expressed in terms of observed variables), which scholars may not be willing to do. On the other hand, to date, very few approaches have been proposed to deal with the issue of nonlinear mediation models using a path-analytic approach, typically used in traditional mediation (Rijnhart et al. 2021, 2023), and they are based either on standardising coefficients (MacKinnon and Dwyer 1993), or on the less employed difference method (Schluchter 2008).

An exception is given by a generalisation of the product method based on partial derivatives, proposed in the ’80 s by Stolzenberg (1980), and recently revived by Hayes and Preacher (2010) and Geldhof et al. (2018). Although quite intuitive, this approach is not widely known and applied by practitioners, and for this reason, it is not well developed, presenting theoretical shortcomings yet to be addressed. The aim of the present paper is to discuss some of these gaps and provide more insights into an approach that has much potential. Specifically, among the insights of the paper, the key ones regard: the implications of employing a binary mediator, how to estimate and interpret the indirect effect when the exposure is discrete, and the proposal of the Bayesian framework for the estimation of confidence intervals for the indirect effects. To the best of our knowledge, these issues have never been addressed in the few previous works discussing the derivative-based approach. As a consequence, the contribution of this paper is two-fold: on the one hand, it tackles several under- or unexplored aspects of the derivative-based approach to nonlinear mediation, proposing novel solutions to some of the issues it presents, and extending it to more complex frameworks like the multilevel one; on the other hand, it

contributes to the existing literature on nonlinear mediation analysis, enriching the underdeveloped group of works in the associational framework.

The paper is structured as follows: first, we briefly review the approaches proposed in the literature to deal with nonlinear mediation models, with a particular focus on the derivative-based approach. Then, in Sect. 3, we introduce our novel contributions to this method, addressing some of its aspects which have received little attention from practitioners, and we believe should be deepened. In the fourth section, we carry out a simulation study addressing some of the issues discussed in Sect. 3, and the fifth section presents an application to real data. The conclusions follow.

2 Mediation analysis in nonlinear models

In this section, we first review the main approaches to address nonlinear mediation models proposed over the years, and then we introduce the derivative-based approach, i.e. the focus of the paper.

2.1 Literature review

In its most basic specification, a mediation model involves three variables: an exposure X, a mediator M and an outcome Y. As already mentioned, in the traditional associational framework, M and Y are continuous and are modelled as linear

$$\mu_M = \beta_0 + \beta_1 X \tag{1}$$

$$\mu_Y = \gamma_0 + \gamma_1 X + \gamma_2 M, \tag{2}$$

where μ_M and μ_Y denote the conditional expectations $\mathbb{E}[M \mid X]$ and $\mathbb{E}[Y \mid X, M]$, respectively.

The coefficient γ_1 represents the direct effect of X on Y, while, as discussed in Baron and Kenny (1986) and Bollen (1989), the indirect effect can be estimated as the product $\beta_1\gamma_2$, i.e., the product of regression coefficients lying on the $X \to M$ and $M \to Y$ paths. Alternatively, considering the marginal model for the outcome, i.e. the model including only the exposure

$$\mu_Y = \alpha_0 + \alpha_1 X, \tag{3}$$

the indirect effect can also be estimated as the difference between α_1, the total effect of X on Y, and the direct effect γ_1, that is $\alpha_1 - \gamma_1$. This approach is called *difference method*, and it is easy to prove that, in the linear case, the product and difference methods yield the same indirect effect estimate (MacKinnon 2008).

In the presence of nonlinearities, such as interaction terms or link functions different from identity, the indirect effect cannot be estimated as a simple product, and its value generally differs from that estimated through the difference method (MacKinnon and Dwyer 1993). The issue of defining and estimating indirect effects in the context of nonlinear models has extensively been addressed in a counterfactual framework (Rubin 2005; Morgan and Winship 2007; Pearl 2009a). Indeed, denoting by $M(x)$ and $Y(x)$ the counterfactual values of the mediator and the outcome if X were set to x, respectively, and by $Y(x, M(x'))$ the counterfactual value of Y if X were set to x and the mediator to

the (natural) value it would have assumed if X were instead equal to $x' \neq x$, Pearl (2001) defined the natural indirect effect (NIE) as

$$NIE = \mathbb{E}[Y(x, M(x)) - Y(x, M(x'))], \tag{4}$$

i.e. the average change in the outcome when leaving the value of X unchanged and moving from the mediator value under the condition $X = x'$ to the mediator value under $X = x$. Under appropriate assumptions about the absence of unobserved confounders (see, for example, Pearl 2001, 2009b; VanderWeele 2015a), it can be proved that the NIE in Eq. (4) can be written as

$$NIE = \sum_{m} \mathbb{E}[Y|x, m]\{P(m|x) - P(m|x')\}. \tag{5}$$

Pearl (2012a, b) highlights how the nonparametric form of Eq. (5) makes its applicability to settings with nonlinear models straightforward. Indeed, he shows how to derive NIE formulas when the mediator and/or the outcome is binary, remarking that the quantities in (5) can be estimated from data directly with no need of fitting parametric models.

Many other counterfactual-based approaches to nonlinear mediation have been proposed in the literature. Valeri and VanderWeele (2013) provide formulas for the NIE in the specific cases of binary and count outcomes, relying on parametric models. They also propose SAS and SPSS macros for the estimation of natural mediational effects. Albert (2012) proposes an estimation approach of natural mediational effects based on a combination of empirical distribution functions and inverse probability weighting, avoiding the need to specify parametric models. Loeys et al. (2013) use a different kind of mediational effects and the so-called natural effects models, highlighting their flexibility in nonlinear contexts. More recently, Gaynor et al. (2019) and Doretti et al. (2022) focused on the setting with a binary mediator and a binary outcome, overcoming the traditional assumption of rareness of the response variable and deriving closed-form expressions for the natural effects based on odds ratios.

In contrast, outside the counterfactual framework, few approaches have been proposed to estimate the indirect effect when the mediator and the outcome are not normal and are modelled with link functions different from identity, for example, using generalised linear models (GLMs). MacKinnon and Dwyer (1993) and MacKinnon (2008) focus on the case of a binary outcome modelled via logistic or a probit model. The indirect effect is obtained as a product of *standardised* regression coefficients. Schluchter (2008) addresses the wider class of GLMs, proposing an extension of the difference method based on generalised estimating equations (GEE). Both these approaches suffer from limitations, the former because it works only for binary outcomes and the latter because it does not allow for exposure-mediator interaction or other forms of nonlinearity in the mediator and the outcome models. It is also worth mentioning the work by Tsai et al. (2006), which extends SEMs to the GLMs framework, but does not discuss how to formalise and estimate indirect effects.

2.2 The derivative-based approach to mediation analysis

The approach discussed in this paper does not rely on counterfactuals and is based on the simple idea that the indirect effect can be interpreted as the variation in the outcome

Y corresponding to a change in the exposure X through the variation in the mediator M (Stolzenberg 1980). Such a definition can be formalised in terms of derivatives, that is:

$$IE = \frac{\partial Y}{\partial M}\frac{\partial M}{\partial X}, \tag{6}$$

i.e. the product of the derivative of Y with respect to M and that of M with respect to X. Let us consider a typical GLM setting with the following models:

$$g_1(\mu_{_M}) = \beta_0 + \beta_1 X \quad \rightarrow \quad \mu_{_M} = h_1(\beta_0 + \beta_1 X) \tag{7}$$

$$g_2(\mu_{_Y}) = \gamma_0 + \gamma_1 X + \gamma_2 M \quad \rightarrow \quad \mu_{_Y} = h_2(\gamma_0 + \gamma_1 X + \gamma_2 M), \tag{8}$$

where g_1 and g_2 are possibly non-linear link functions, connecting the conditional expectations of the mediator and the outcome to their linear predictors, and $h_k = g_k^{-1}$, $k \in \{1, 2\}$.

Notice that in the trivial case of identity link functions, the indirect effect in formula (6) reduces to the traditional expression obtained via the product method $\beta_1\gamma_2$.

In contrast, when at least one of the g functions differs from identity, the indirect effect assumes a more complex form. In this case, the expression of the indirect effect is not a single value but depends on X and/or M via the derivatives of h_1 and h_2. Assuming a continuous exposure, the researcher chooses some values of X of potential interest, say $x_1^*, \ldots, x_p^*$ and, if the expression of the indirect effect also involves the mediator, its values should be selected accordingly to those of X, as the predicted values corresponding to $x_1^*, \ldots, x_p^*$, obtained from the fitted model $\mu_{_{M_{|x_1^*}}}, \ldots, \mu_{_{M_{|x_p^*}}}$. For this reason, Geldhof et al. (2018) suggest calling the effect in Equation (6) *Conditional Indirect Effect* (CIE), since its values are conditional to those of X. To illustrate this concept further, let us consider the case where both the mediator and the outcome are binary variables. Let us also assume that they are both modelled using logistic regression. In this case the indirect effect relative to x_p^*, using the model specification in Equations (7) and (8) is found as:

$$CIE_{|x_p^*} = \frac{\beta_1 \cdot \exp(\beta_0 + \beta_1 \cdot x_p^*)}{(1 + \exp(\beta_0 + \beta_1 \cdot x_p^*))^2} \cdot \frac{\gamma_2 \cdot \exp(\gamma_0 + \gamma_1 \cdot x_p^* + \gamma_2 \cdot \mu_{_{M_{|x_p^*}}})}{(1 + \exp(\gamma_0 + \gamma_1 \cdot x_p^* + \gamma_2 \cdot \mu_{_{M_{|x_p^*}}}))^2}.$$

As highlighted above, in the presence of these nonlinearities, the Indirect Effect becomes dependent on the values of x_p^* and $\mu_{_{M_{|x_p^*}}}$, and this explains why the effect was named 'conditional' by Geldhof et al. (2018).

The models in Eqs. (7)–(8) are intentionally very simple, but real-world data generally require adjustment for covariates. Including covariates $\boldsymbol{Z}$ in the mediator and the outcome models affects the expression of the indirect effect, which may depend on the covariates' values in addition to those of X and M. For example, consider models as in Eqs. (7)–(8), where h_1 is the identity (i.e. the mediator model is linear), and h_2 is the exponential function, and include two (possibly overlapping) sets of covariates $\boldsymbol{Z}_M$ and $\boldsymbol{Z}_Y$ for the mediator and the outcome, respectively:

$$\mu_{_M} = \beta_0 + \beta_1 X + \sum_{k=1}^{p} \beta_{k+1} Z_{Mk}$$

$$\mu_{_Y} = \exp\left(\gamma_0 + \gamma_1 X + \gamma_2 M + \sum_{k=1}^{q} \gamma_{k+2} Z_{Yk}\right).$$

The indirect effect is

$$CIE = \beta_1 \gamma_2 \exp\left(\gamma_0 + \gamma_1 X + \gamma_2 M + \sum_{k=1}^{q} \gamma_{k+2} Z_{Yk}\right)$$

i.e., substituting $\mu_{_M}$ to M,

$$CIE = \beta_1 \gamma_2 \exp\left(\gamma_0 + \gamma_1 X + \gamma_2 \left(\beta_0 + \beta_1 X + \sum_{k=1}^{p} \beta_{k+1} Z_{Mk}\right) + \sum_{k=1}^{q} \gamma_{k+2} Z_{Yk}\right)$$

As already mentioned, the formula also includes covariates. Hayes and Preacher (2010) suggest estimating the indirect effects conditional on the values of X and M, setting the covariates to their mean values. The authors do not address the scenario where the covariates act as effect modifiers, i.e. when they interact with the exposure or the mediator. This simply makes the partial derivatives in Eq. (6) more complex but does not add any conceptual difficulty. An important interaction term, which is often included in the outcome model, is that between the exposure and the mediator. To see how the presence of such a term influences the indirect effect, it is sufficient to consider linear models for both the mediator and the outcome and include a term $\gamma_3 XM$ in the outcome model. The indirect effect is $\beta_1(\gamma_2 + \gamma_3 X)$, which depends on X, in contrast to the indirect effect obtained from models excluding the presence of an exposure-mediator interaction, i.e. the simple product $\beta_1 \gamma_2$. The expression $\beta_1(\gamma_2 + \gamma_3 X)$ is consistent with that obtained by VanderWeele (2015b) in a counterfactual-based framework.

The main focus of this paper is on mediation analysis with GLMs; nonetheless, it is worth remarking that the derivative-based approach can also be used in situations where the mediator or the outcome depends on nonlinear transformations of their regressors, such as X^2 or $\log(X)$, see Hayes and Preacher (2010) for some examples. Moreover, this approach is crucial even when the distributions of variables M and Y do not belong to the exponential family. In such cases, the key is establishing appropriate link functions that enable us to calculate the derivatives and obtain the CIE.

3 Insights and extensions of the derivative-based mediation approach

In this section, we discuss some relevant aspects of the derivative-based method which have not been addressed or satisfactorily deepened in the previous literature on the topic. For the sake of generality, we assume to be in an observational setting, although moving to an experimental setting will not influence the results.

3.1 Binary, categorical and discrete exposures

The mathematical definition of derivatives is based on the concept of 'small increment' in the argument of the function to differentiate. Thus, the derivative of the mediator with respect to the exposure conceptually relies on a small increment of X, and analogously for the derivative of the outcome with respect to the mediator. This definition makes sense if the support $\mathcal{D}_W$ of the variable W with respect to which differentiation is made is continuous. The derivative is a continuous function defined over $\mathcal{D}_W$ or a subset. However, when W is discrete, the interpretation of the derivative becomes more challenging, as that of the indirect effect in Eq. (6).

Let us start with a setting with a binary exposure. The mediator expectation is then a discrete function that can assume only two values. Consequently, the concept of infinitesimal increment is misspecified since the only increment meaningful to conceive is a unit increment. Derivatives cannot be applied to discrete functions; therefore, an alternative definition to express change is required. We can use finite differences

$$D_{x,w}[f] = \frac{f(x+w) - f(x)}{w} \tag{9}$$

where f is the function of interest and w is the difference between two points in $\mathcal{D}_X$. Notice that, when $w \to 0, D_{x,w}(f) \equiv \frac{df}{dx}$. Going back to our mediational setting, the derivative of the mediator in the case of binary exposure can then be written simply as the difference

$$D_{0,1}[h_1] = h_1(1) - h_1(0). \tag{10}$$

It is easy to prove (see "Appendix") that the chain rule for composite functions holds also in the discrete case, and the indirect effect can then be written as

$$CIE = D_{\mu_M(x),wD_{x,w}[\mu_M(x)]}[\mu_Y(\mu_M(x))] \cdot D_{x,w}[\mu_M(x)],$$

where we explicitly wrote the functional dependence of μ_M and μ_Y, and where the variation of x concerns only μ_M.

The case of binary exposure easily extends to that of categorical exposure. Suppose that X is a categorical variable with k categories. Without loss of generality, assume that the k-th category is the one chosen as baseline. The mediation model can be rewritten as

$$h_1(\mu_M) = \beta_0 + \sum_{j=1}^{k-1} \beta_j X_j, \tag{11}$$

where the X_j are binary variables assuming value 1 if X is in category j, 0 otherwise. In this case, it is necessary to specify the variable with respect to which one takes the difference or, in other words, the category with respect to which the indirect effect is estimated. An increment from 0 to 1 represents the passage from the baseline to this selected category.

The same line of reasoning holds for discrete exposures, for example, number of cigarettes smoked in a day or number of panic attacks in a month.

3.2 Binary mediator

As already mentioned, when the expression of the indirect effect involves both X and M, the values of M cannot be chosen arbitrarily, instead they should be fixed at the values

the mediator takes in correspondence of the selected values of X, determined by the fitted model. This is generally straightforward unless the mediator is binary when some issues arise.

Consider a setting with a binary mediator, where g_1 in Eq. (7) is the logit link and, and g_2 in the outcome model in Eq. (8) is a generic function different from identity, say the logarithm to fix ideas. Therefore, applying the formula in Eq. (6), the CIE is given by

$$CIE = \beta_1 \gamma_2 \frac{\exp(\beta_0 + \beta_1 X)}{(1 + \exp(\beta_0 + \beta_1 X))^2} \exp(\gamma_0 + \gamma_1 X + \gamma_2 M),$$

which, as can be seen, depends on both X and M. However, when coming to the estimation of such an effect, the choice of the mediator value is not so immediate. Indeed, the mediator is binary, assuming only two values, while, for any selected value of X, its predicted values from model (7) with a logit link are probabilities, ranging in the continuum from 0 to 1. Which values to select, then? This issue is addressed neither by Hayes and Preacher (2010) nor by Geldhof et al. (2018). We believe that the most appropriate solution consistent with a data-generating mechanism of this type is to include binary values of the mediator obtained by the corresponding expected probabilities $\hat{\pi}_{M|X}$ by means of a cutoff c such that

$$\hat{M}|X = \begin{cases} 1 & \text{if } \hat{\pi}_{M|X} \geq c \\ 0 & \text{if } \hat{\pi}_{M|X} < c \end{cases}$$

The most trivial choice $c = 0.5$ may often not be appropriate, for example when classes are unbalanced. A possible alternative criterion for the choice of c could be selecting the value for which the sensitivity and specificity of the classification are equal based on the ROC curve, which can ensure a better performance. Clearly, other approaches are possible, for example maximizing the Youden's index or the F1-score (Berrar 2019).

In principle, this idea can be extended to the case of categorical mediators, for which it is necessary to estimate multiple thresholds. There exist generalisations of the ROC curve for polytomous variables; see Hand and Till (2001) for example. However, moving from a setting with only two categories to a setting with multiple categories comes at the cost of additional issues. These issues go beyond the scope of this paper, suggesting the need for further investigation of this topic.

3.3 Extension to multilevel models

Geldhof et al. (2018) claim that the derivative-based approach can be easily extended to the multilevel case, but we are not aware of any study addressing this issue. In the following, we discuss how the derivative-based method can be applied to generalized mixed-effect models (GLMMs).

Consider a setting with J clusters and $I = \sum_j n_j$ subjects, where n_j is the number of individuals belonging to each cluster. Typical examples of clustered data are children within classrooms, employees in an organization's departments, or patients in hospitals. Let us start from linear multilevel models where all variables are measured at the subject level (level 2), i.e. a $1 \rightarrow 1 \rightarrow 1$ design, using the notation introduced by Krull and MacKinnon (1999, 2001):

$$\mu_{M_{ij}} = (\beta_0 + b_{0j}) + (\beta_1 + b_{1j})X_{ij} \tag{12}$$

$$\mu_{Y_{ij}} = (\gamma_0 + g_{0j}) + (\gamma_1 + g_{1j})X_{ij} + (\gamma_2 + g_{2j})M_{ij} \tag{13}$$

where j denotes the cluster and i the subject, Greek letters denote fixed effects and $(b_{0j}, b_{1j}, g_{0j}, g_{1j}, g_{2j})'$ is the vector of random effects, which are assumed to be from a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}$ possibly non-diagonal (i.e. random effects are allowed to have non-null covariances). Random effects capture the interdependence of units belonging to the same cluster and represent cluster-specific deviances from the average intercept or slope levels.

Using Equation (6), the indirect effect is the product:

$$IE = (\beta_1 + b_{1j})(\gamma_2 + g_{2j}), \tag{14}$$

which depends on b_{1j} and g_{2j}, for $j = 1, \ldots, J$, or, in other words, the indirect effect is cluster-specific. To obtain a unique, average indirect effect, it is necessary to integrate random effects out. If b_1 and g_2 are uncorrelated, then the indirect effect is simply the product $\beta_1\gamma_2$; when, however, they are correlated, then the indirect effect is given by

$$IE = \beta_1\gamma_2 + \sigma_{b_1 g_2}, \tag{15}$$

where $\sigma_{b_1 g_2}$ is the covariance between b_1 and g_2 (Kenny et al. 2003). The estimation of such a covariance term in the traditional multilevel setting is complex and requires *ad hoc* solutions, like those proposed in Kenny et al. (2003) and Bauer et al. (2006), or to address multilevel models from a structural perspective (Bauer et al. 2006; Curran 2003; Preacher et al. 2010, 2011). Another option could be moving to a Bayesian framework (Yuan and MacKinnon 2009; Di Maria et al. 2022), which allows us to obtain the posterior distribution of $\sigma_{b_1 g_2}$ and of the indirect effect, also making the estimation of confidence intervals straightforward.

When at least one between h_1 and h_2 differs from identity, or, in other words, when the mediator and/or the outcome model is a generalized mixed model, estimation becomes more complex. For example, if the mediator is a count variable and we model it using a log link,

$$\mu_{M_{ij}} = \exp\{(\beta_0 + b_{0j}) + (\beta_1 + b_{1j})X_{ij}\}$$

while the outcome follows a linear model as in Eq. 13, the indirect effect is

$$CIE = (\beta_1 + b_{1j})(\gamma_2 + g_{2j}) \exp\{(\beta_0 + b_{0j}) + (\beta_1 + b_{1j})X_{ij}\},$$

for which integrating out random effects and obtaining a closed form expression may be complex or even not feasible. In this case, numerical integration methods may be necessary.

So far, we have focused on the $1 \rightarrow 1 \rightarrow 1$ design, the one generally most complex due to the potential presence of a product between two random effects. Analogous considerations can, however, be extended to other multilevel mediational designs, like $2 \rightarrow 1 \rightarrow 1$ or $2 \rightarrow 2 \rightarrow 1$[1]. When both the mediator and the outcome models are linear, the integration of

[1] It is important to remark that, according to Krull and MacKinnon (1999, 2001), these are the only designs that can be addressed in the traditional multilevel framework. Other designs, including $1 \rightarrow 2$ components, i.e. *bottom-up* effects, cannot be dealt with.

random effects is straightforward, while when one or both models are nonlinear, some difficulties may arise, especially in the case of correlated random effects.

3.4 Confidence intervals for the indirect effect

Estimating confidence intervals (CIs) for the indirect effect poses challenges even in the linear case, since the distribution of the product $\beta_1\gamma_2$ does not follow a Normal distribution although the two coefficient estimators are assumed to be Normal (Springer and Thompson 1966; Lomnicki 1967). Indeed, the distribution of the product may be asymmetric and difficult to approximate with distributions traditionally used in statistics, see MacKinnon (2008). This issue is potentially exacerbated in a nonlinear setting, where the indirect effect assumes complex forms, the distribution of which may be impractical (or simply impossible) to derive in closed form. Therefore, it seems convenient to rely on sampling-based approaches to retrieve an empirical distribution of the indirect effect, from which to compute statistics of interest. Geldhof et al. (2018) suggest using non-parametric bootstrap or Monte-Carlo confidence intervals (for a reference see, e.g., Efron and Tibshirani (1994) and Rubinstein and Kroese (2016)). The former creates B samples by resampling statistical units in the original sample, and for each of them, the parameter of interest is estimated, the indirect effect in our case.

The latter method does not require data resampling, but it generates samples of regression parameters in (7)–(8), assuming that they come from a multivariate normal distribution. Each of these samples is associated to an estimate of the indirect effect or, more generally, to the parameter of interest.

Supported by a growing body of literature which highlights its desirable properties (see, for example Yuan and MacKinnon 2009; Biesanz et al. 2010; Koopman et al. 2015; Miočević et al. 2017), we believe that another valuable option for the estimation of CIs could be the Bayesian approach (for an introduction see, e.g., Gelman et al. (2013)). Each parameter is endowed with an *a priori* distribution, and an empirical (posterior) distribution of the indirect effect can be obtained via one of several methods available for this purpose, e.g. Monte Carlo Markov Chains (MCMC). Moving to the Bayesian framework can also provide additional advantages, like the possibility to embed prior information into the mediation model, if available, in order to improve estimates efficiency, ease of extension to the multilevel case, even assuming complex *a priori* correlation structures of fixed and random effects, and exact inference for small samples, for which asymptotic assumptions might not hold.

To the best of our knowledge, no simulation studies have been run so far to compare the performance of these three approaches for indirect effects estimated through the derivative-based method. This is the primary focus of Sect. 4.

4 Simulation study

In order to overcome the issues related to the closed-form estimation of confidence intervals for indirect effects in a non-linear context, sampling-based approaches may be a possible alternative. In particular, we focused on Bootstrap, Monte Carlo and Bayesian intervals,

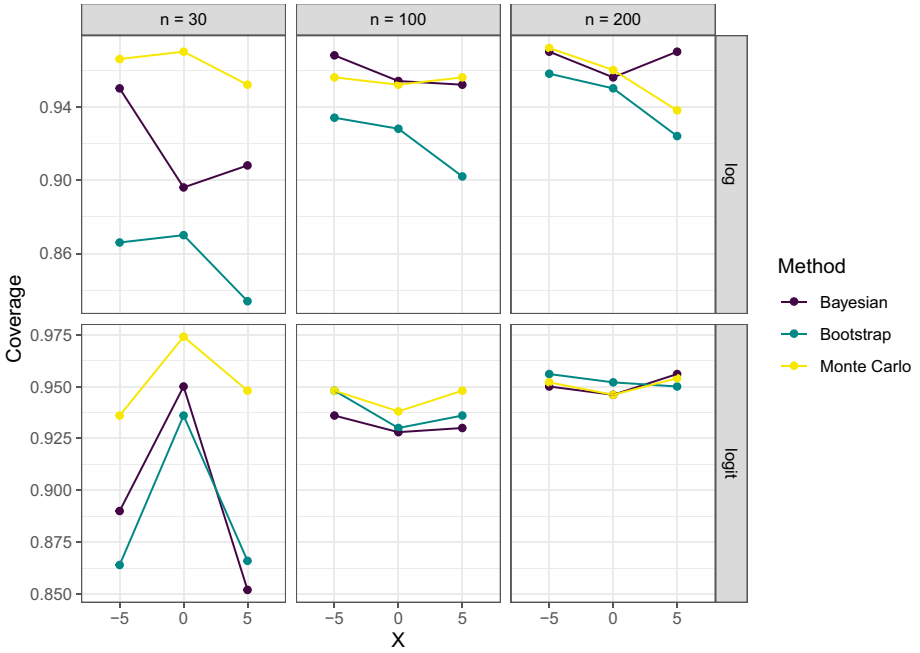

Fig. 1 Results of the simulation study: coverage rates

conducting a simulation study in order to compare their behaviour under different conditions. Namely, we considered three sample sizes ($n = 30, 100, 200$) and two combinations of assumed distribution and link function for both the mediator and the outcome models, that is Bernoulli distribution and logit link in one case and Poisson distribution and log link in the other, with expectations as in Eqs. (7)–(8). We generated the exposure variable X from a $\mathcal{N}(0, 5^2)$, and we chose three different exposure values (0 and ± 5) on which to condition the indirect effect. The values of the model coefficients were chosen arbitrarily.

Bootstrap estimates are obtained by drawing 1,000 samples of size n, with replacement, from the original simulated dataset. Then, with each bootstrap sample, we fit models in Eqs. (7) and (8) and at each iteration we saved coefficients' estimates. To retrieve Monte Carlo samples, we first estimated GLM expressed in Eqs. (7) and (8), then we sampled 1000 regression coefficients values from a $\mathcal{MVN}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}}_{\hat{\theta}})$, where

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \end{bmatrix} \quad \hat{\boldsymbol{\Sigma}}_{\hat{\theta}} = \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_{\hat{\beta}} & \mathbf{0} \\ \mathbf{0} & \hat{\boldsymbol{\Sigma}}_{\hat{\gamma}} \end{bmatrix} \tag{16}$$

$\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ being the vectors of estimated coefficients of mediator and outcome models, respectively, and $\hat{\boldsymbol{\Sigma}}_{\hat{\beta}}$ and $\hat{\boldsymbol{\Sigma}}_{\hat{\gamma}}$ their asymptotic estimated covariance matrices. Matrix $\hat{\boldsymbol{\Sigma}}_{\hat{\theta}}$ is block diagonal because we assume that $Cov(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) = 0$. Bayesian posterior coefficients samples have been derived using diffuse priors ($\mathcal{N}(0, 10^3)$) for each parameter, by means of Monte Carlo Markov Chains, from two chains of length 10,000 with burn-in = 5000. Graphical inspection of the chains showed that all the chains converged. Simulations were carried out in `R`, using the package `rjags` for the bayesian part.

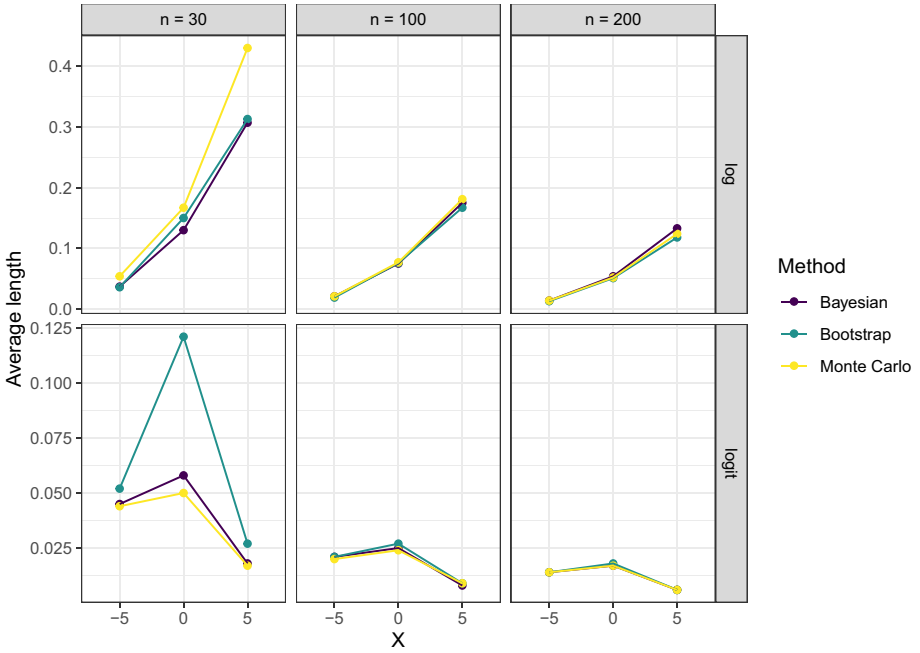

Fig. 2 Results of the simulation study: average CI lengths

We repeated the process 500 times, each time estimating the value of the indirect effect, in order to retrieve its empirical distribution for each approach and scenario, then we computed quantile-based 95% intervals. We compared the three approaches in terms of the average length of the intervals and the proportion of intervals which contain the "true" value of the indirect effect (i.e. coverage rate). Results are shown in Figs. 1 and 2 and in Table 1.

As expected, the average CI lengths and the differences between the three methods reduce as the sample size increases. When $n = 30$, in general, it can be noticed that the coverage rate of Monte Carlo CIs is slightly higher than that of the other two approaches. Bayesian intervals behave well compared to the others and can then be considered as a reasonably good alternative.

5 Application

In this section, we analyse data from the ANS (Anagrafe Nazionale Studenti), which serves as the database for Italian university students (MOBYSU.IT. 2017). Each record in the database represents a statistical unit, specifically a freshman enrolled at an Italian university. These records contain variables about the student's high school background and university career. For this study, we have chosen to focus on the 2015 cohort, the most recent available cohort, which covers a sufficiently long time span to observe the completion of the degree. We have decided to limit the analysis to students enrolled at a

Table 1 Results of the simulation study. For each method, the average coverage rates and interval lengths are reported

Sample size	Link	X value	True eff.	Bayesian		Bootstrap		Monte Carlo	
				Cov. rate	Avg. length	Cov. rate	Avg. length	Cov. rate	Avg. length
30	logit	−5	−0.005	0.890	0.045	0.864	0.052	0.936	0.044
		0	−0.007	0.950	0.058	0.936	0.121	0.974	0.050
		5	−0.002	0.852	0.018	0.866	0.027	0.948	0.017
	log	−5	−0.010	0.950	0.037	0.866	0.036	0.966	0.054
		0	−0.066	0.896	0.130	0.870	0.150	0.970	0.167
		5	−0.097	0.908	0.307	0.834	0.313	0.952	0.403
100	logit	−5	−0.005	0.936	0.021	0.948	0.021	0.948	0.020
		0	−0.007	0.928	0.025	0.930	0.027	0.938	0.024
		5	−0.002	0.930	0.008	0.936	0.009	0.948	0.009
	log	−5	−0.012	0.968	0.021	0.934	0.019	0.956	0.021
		0	−0.097	0.954	0.075	0.928	0.076	0.952	0.077
		5	[−0.133]	0.952	0.175	0.902	0.167	0.956	0.181
200	logit	−5	−0.005	0.950	0.014	0.956	0.014	0.952	0.014
		0	−0.007	0.946	0.017	0.952	0.018	0.946	0.017
		5	−0.002	0.956	0.006	0.950	0.006	0.954	0.006
	log	−5	−0.012	0.970	0.014	0.958	0.013	0.972	0.014
		0	−0.095	0.956	0.054	0.950	0.051	0.960	0.052
		5	−0.124	0.970	0.133	0.924	0.118	0.938	0.124

non-online Sicilian university, comprising $N = 19\,770$ individuals. It is worth emphasising that the exclusion of students enrolled at online universities is driven by their unique behavioural patterns concerning degree completion (Priulla 2023).

Our analysis employs associational nonlinear mediation analysis to examine the relationship between high school background and academic success. Specifically, the goal is to examine the impact of the high school final mark (HSFM) on the probability of achieving a bachelor's degree (BD) within four academic years while also exploring the mediating role of the number of University Credits (UC) earned at the end of the first year.

In addition to the HSFM, CU and BD variables, which act as exposure, mediator and outcome, respectively, we included a set of covariates to control for possible confounders. The three variables of interest and the selected covariates are briefly described below:

- `HSFM`: final mark obtained by the student at the end of high school. In Italy, the final mark ranges from 60 ('Sufficient') to 100 *cum laude*, coded as 101. Decimal scores are not allowed.
- `UC`: number of university credits obtained by the student within the first year from his enrollment to their current degree course. Generally, the maximum number of credits a student can get during the first year is 60.

- `BD`: binary variable, taking value 1 if the student graduates within four years from their enrolment to the current degree course, 0 otherwise.
- `TUSS`: Type of Upper Secondary School diploma. In Italy, there are various types of upper secondary schools, each offering a different curriculum and training students for a particular career or academic path. In this study they have been categorized in Classical *lyceum*, Scientific *lyceum*, Technical institute, Industrial Technical institute, Vocational institute, Industrial Technical institute, Other *lyceum* (baseline), and Abroad/Other.
- `sex`: student's biological sex, male and female (baseline).
- `TDC`: area of the degree course at which the student is enrolled, categorised in "Agriculture, forestry, fisheries and veterinary" (baseline), "Arts and humanities", "Engineering, manufacturing and construction", "Health and welfare", "Business, administration and law", "Natural sciences, mathematics and statistics (NsMS)", "Services", "Social sciences, journalism and information", "Education", and "Information and Communication Technologies (ICTs)".
- `age`: student's age.

In this case study, the variables UC and BD acting as the mediator and outcome, respectively, do not follow a Gaussian distribution, suggesting the need for nonlinear mediation analysis. Specifically, the UC variable is bounded between 0 and 60, as mentioned before. To account for this, we first transformed UC into the proportion of UC (PUC) obtained in the first year, dividing UC by 60; then, to make these scores strictly in the interval (0,1), we applied the transformation proposed by Smithson and Verkuilen (2006):

$$PUC' = \frac{PUC \cdot (N-1) + 0.5}{N},$$

and employed a Beta model with logit link for analysis:

$$\text{logit}\big(\mathbb{E}[PUC'|X,Z]\big) = \beta_0 + \beta_1 \text{HSFM} + \boldsymbol{\beta_2} \text{TUSS} + \beta_3 \text{sex} + \boldsymbol{\beta_4} \text{TDC} + \beta_5 \text{age}.$$

Regarding the outcome, BD is a binary variable that takes the value of 1 if the student successfully graduates on time. To examine the relationship between HSFM and PUC$'$, we employed a logistic regression model as follows:

$$\text{logit}(P[BD = 1|M,X,Z]) = \gamma_0 + \gamma_1 \text{HSFM} + \gamma_2 \text{PUC}' + \boldsymbol{\gamma_3} \text{TUSS} + \gamma_4 \text{sex} + \boldsymbol{\gamma_5} \text{TDC} + \gamma_6 \text{age}.$$

An important point to highlight is that the exposure variable, HSFM, is a discrete variable. Consequently, estimating the indirect effect requires the use of finite differences methodology. Since the model also includes some other covariates, they need to be fixed to specific values. Specifically, we fixed TUSS, TDC, and age to their joint mode: TUSS = Scientific *lyceum*, TDC = Engineering, manufacturing and construction, and age = 19. In contrast, the variable sex has not been explicitly assigned. Indeed, we calculated the CIEs for both males and females, enabling a meaningful comparison. The indirect effect for the i-th value of HSFM is found as:

$$\text{CIE}_i = \text{logit}^{-1}(\eta_{i+1}) - \text{logit}^{-1}(\eta_i), \quad i = 1, \dots, 41, \tag{17}$$

where

Table 2 Estimates and p values of regression coefficients involved in the estimation of CIEs

	Mediator			Outcome		
Name	Coef	Estimate	p value	Coef	Estimate	p value
Intercept	β_0	−2.683	< 2e-16	γ_0	−2.900	< 2e-16
HSFM	β_1	0.039	< 2e-16	γ_1	0.0004	0.774
PUC′	-	-	-	γ_2	4.583	< 2e-16
TUSS = Sci	β_2	0.361	< 2e-16	γ_3	−0.050	0.345
sex = male	β_3	−0.032	0.124	γ_4	−0.039	0.321
TDC = Engineering	β_4	−0.114	0.015	γ_5	−0.302	0.001
age	β_5	−0.034	< 2e-16	γ_6	−0.001	0.903

$$\eta_i = \gamma_0 + \gamma_1 \text{HSFM}_i + \gamma_2 \text{PUC}'_{|HSFM_i} + \gamma_3^{(TUSS=Sci)} + \gamma_4 \text{sex} + \gamma_5^{(TDC=Eng)} + \gamma_6 \cdot 19$$

and

$$PUC'_{|HSFM_i} = \text{logit}^{-1}(\beta_0 + \beta_1 \text{HSFM}_i + \beta_2^{(TUSS=Sci)} + \beta_3 \text{sex} + \beta_4^{(TDC=Eng)} + \beta_5 \cdot 19).$$

Note that $\text{HSFM}_1 = 60$ and $\text{HSFM}_{42} = 101$. In this terms, CIE_i quantifies how the probability of achieving BD changes when HSFM increases by one unit, from HSFM_i to HSFM_{i+1} in the mediator model, leaving its value fixed in the outcome model, considering the mediating effect of PUC′.

The coefficients involved in the estimation of the CIEs are reported in Table 2, while the whole set of coefficients can be found in the Appendix (see Table 3).

The results in Table 2 seem to suggest that the relationship between HSFM and BD is fully mediated by PUC′ since the effect of HSFM in the outcome model, γ_1, is not significant. HSFM is positively and significantly associated with PUC′, which is in turn positively and significantly associated with BD. The magnitude of the latter coefficient is remarkable (4.583). To formally test if the indirect effect is significant, we estimated the CIEs (as shown in Eq. 17) and their confidence intervals using the three approaches discussed before.

Figure 3 shows the results obtained using the Bayes approach, while those obtained with the other approaches are almost identical (see Figs. 4 and 5 in the Appendix); actually, this is in agreement with what we observed in simulations in the scenario with a large sample size, as it is here.

Each point in the graph represents the difference in the probability of graduating within four years for a unitary increase in HSFM mediated by PUC′. It is worth noting that all the estimated CIEs are positive and significant (their CIs does not contain zero), meaning that getting a higher HSFM is associated with a higher probability of graduating on time through PUC′. However, the curve has a monotonic increasing trend until HSFM reaches 92, then it slightly starts to decrease. This may suggest that the mediating role of PUC′ becomes more and more important as HSFM increases until 92, at which point it becomes slightly less relevant. In addition, we can notice that the indirect effect for females is slightly larger than those of males; however, the confidence intervals overlap for all values of HSFM, implying that the observed differences in CIEs magnitude are not significant. This is consistent with the regression model's results.

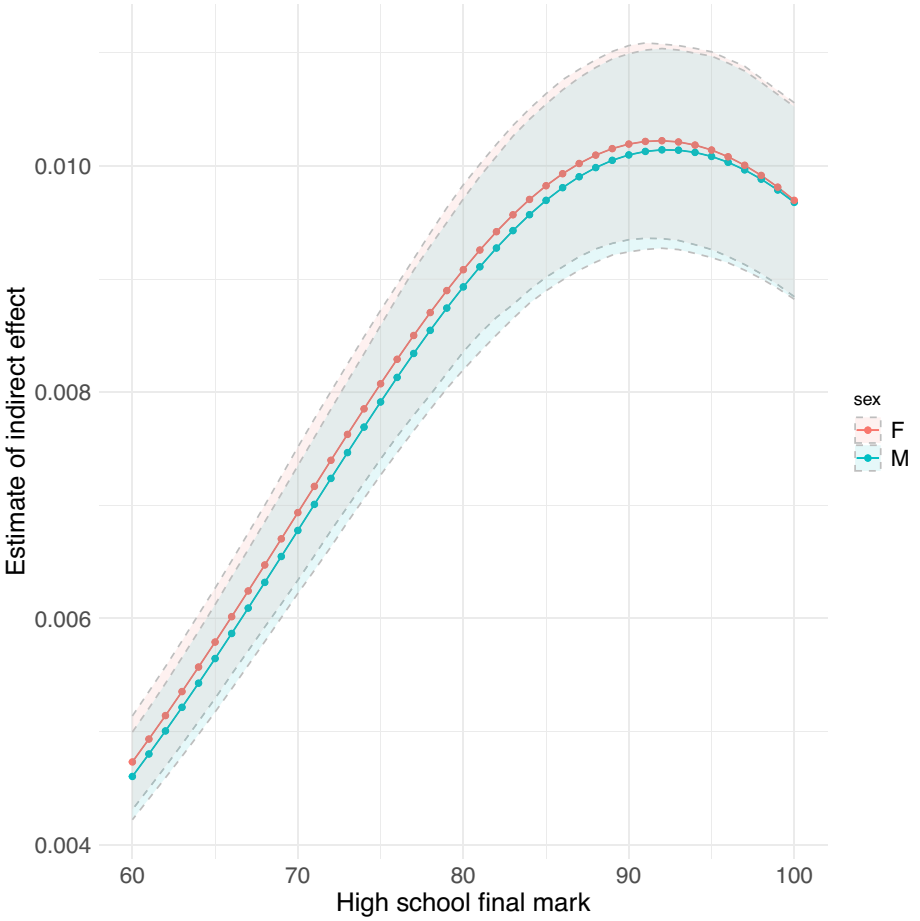

Fig. 3 CIEs for males and females and their confidence intervals estimated with the Bayesian approach

6 Conclusions

Estimating the indirect effect in an associational mediational context where the mediator and/or the outcome model are nonlinear is a common and relevant task in applied research but is not straightforward. The existing literature primarily focuses on nonlinear mediation models within the causal framework, which requires different assumptions and notation compared to the traditional framework commonly used in mediation analysis (Rijnhart et al. 2021, 2023). Stolzenberg (1980) proposed to estimate the indirect effect as the derivative of a composite function, and Geldhof et al. (2018) discussed some applications of this approach in the GLM context. In this work, we offer a comprehensive discussion of the derivative-based method for nonlinear mediation analysis by deepening some aspects of the proposal of Geldhof et al. (2018), addressing potential issues, like the presence of binary mediators and the corresponding choice of predicted values for the estimation of CIEs, and how to deal with binary/categorical/discrete exposure variables. We also proposed the use of a Bayesian approach as a valuable option for the estimation of CIEs

confidence intervals, which also allows researchers to include *a priori* information about variables, offers exact inference for small samples, for which asymptotic assumptions might not hold, and can provide estimates for the covariance between random effects in multilevel mediation models. Through a simulation study, we demonstrated that Bayesian intervals are a valid alternative compared to the Bootstrap-based and Montecarlo ones used by Geldhof et al. (2018).

Also, we present a real-data application which investigated the relationship between high-school background and academic success, in which we dealt with the presence of a discrete exposure variable. This posed a limitation for applying the derivative-based approach discussed in the paper, as the concept of derivative lacks meaning in the context of discrete variables. To overcome this challenge, we employed finite differences to estimate the conditional indirect effects. By calculating finite differences, we were able to capture the effects of the mediator on the outcome variable while accounting for the discreteness of the high-school background variable. The results obtained through this approach were then interpreted accordingly, acknowledging the specific characteristics of the variables involved.

Overall, we believe that our work can contribute to the existing literature on nonlinear mediation analysis by discussing a very promising approach and addressing some of its potential issues through novel solutions. In addition, this paper can serve as a guide for researchers who need to address a mediational setting with nonlinear models without switching to the counterfactual framework. Our work can be extended in several ways. First, settings with categorical mediators need further investigation. This issue has traditionally received little attention by scholars, either in the counterfactual and the associational framework, and can be a promising research direction. Another interesting issue concerns the relationship between the total and indirect effects. Indeed, in the classical linear setting, the total effect is the sum of the direct and indirect effects; however, this property does not hold in the case of nonlinear models. Another possible venue for future research is clustered data, for which multilevel models are often employed. A simulation study may show the strengths and limits of the three approaches for estimating confidence intervals illustrated in the previous sections in such a setting. This underscores the need for continued exploration and development of methodologies for nonlinear mediation analysis in diverse real-world scenarios.

Appendix A: Chain rule for finite differences

Let $f(x)$ and $g(x)$ two discrete functions and $f \circ g \equiv f(g(x))$ the function obtained from their composition. We want to prove that

$$D_{x,w}[f(g)] = \frac{f(g(x+w)) - f(g(x))}{w}$$

can be written as a product of differences, deriving a chain rule analogous to that for derivatives of continuous functions. Indeed, $D_{x,w}[f(g)]$ can be written as

$$\begin{aligned}\frac{f(g(x+w)) - f(g(x))}{w} &= \frac{f\left(g(x) + w\,\frac{g(x+w)-g(x)}{w}\right) - f(g(x))}{w} \\ &= \frac{f\left(g(x) + wD_{x,w}[g]\right) - f(g(x))}{w}\end{aligned}$$

Table 3 Estimates and p values of regression coefficients

	Mediator			Outcome		
Name	Coef	Estimate	p value	Coef	Estimate	p value
Intercept	β_0	− 2.683	< 2e−16	γ_0	− 2.900	< 2e−16
HSFM	β_1	0.039	< 2e−16	γ_1	0.0004	0.774
PUC′	–	–	–	γ_2	4.583	< 2e−16
TUSS = Sci	β_2	0.361	< 2e−16	γ_3	− 0.050	0.345
TUSS = Clas		0.209	< 2e−16		− 0.301	< 2e−16
TUSS = Tech		− 0.209	< 2e−16		− 0.137	0.039
TUSS = Voc		− 0.38	< 2e−16		− 0.338	0.002
TUSS = Ind Tech		− 0.101	0.047		− 0.066	0.525
TUSS = Abroad		0.021	0.894		0.189	0.54
sex = male	β_3	− 0.032	0.124	γ_4	− 0.039	0.321
TDC = Engineering	β_4	− 0.114	0.015	γ_5	− 0.302	0.001
TDC = Arts		0.438	< 2e−16		0.092	0.331
TDC = Health		0.326	< 2e−16		− 0.555	< 2e−16
TDC = Business		− 0.017	0.708		− 0.865	< 2e−16
TDC = NsMS		− 0.425	< 2e−16		0.453	< 2e−16
TDC= Education		0.734	< 2e−16		− 0.921	< 2e−16
TDC = ICTs		− 0.299	<0.001		− 0.141	0.419
age	β_5	− 0.034	< 2e−16	γ_6	− 0.001	0.903

Noting that

$$D_{g(x),wD_{x,w}[g]}[f(g)] = \frac{f(g(x) + wD_{x,w}[g]) - f(g(x))}{wD_{x,w}[g]},$$

it is easy to derive that

$$D_{x,w}[f(g)] = D_{g(x),wD_{x,w}[g]}[f(g)] \cdot D_{x,w}[g].$$

Appendix B: Real data application supplementary

B.1 Regression coefficients

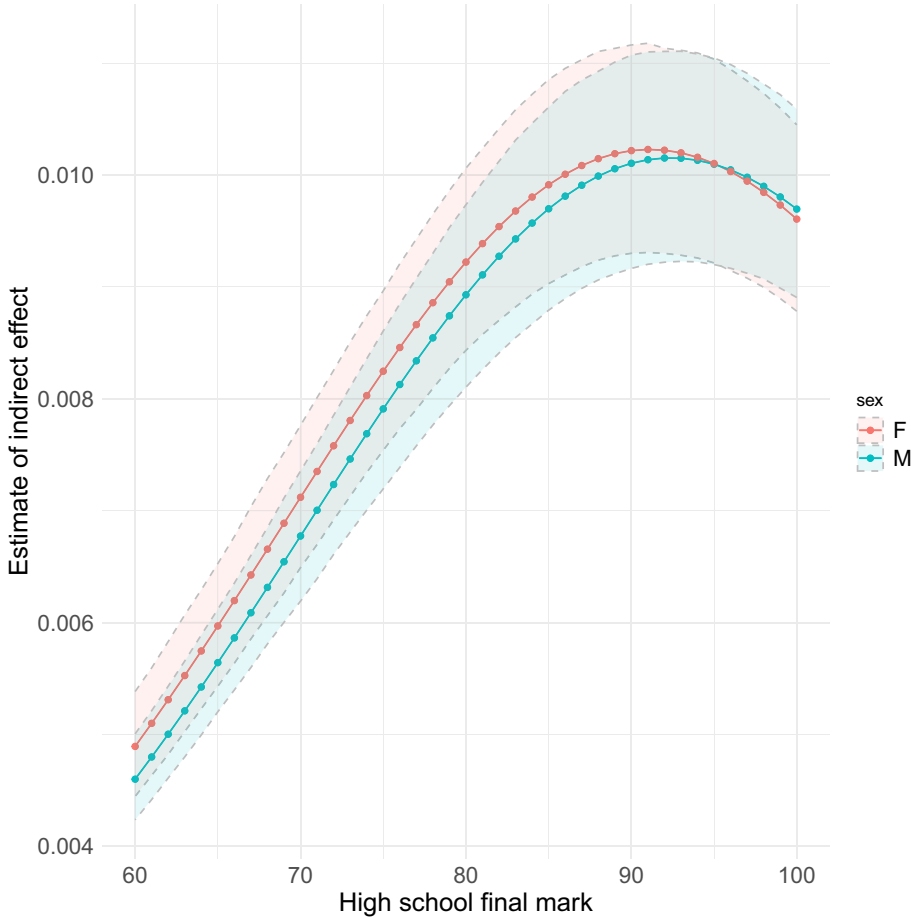

Fig. 4 CIEs for males and females and their confidence intervals estimated with the Bootstrap approach

B.2 Estimates of CIEs on real data

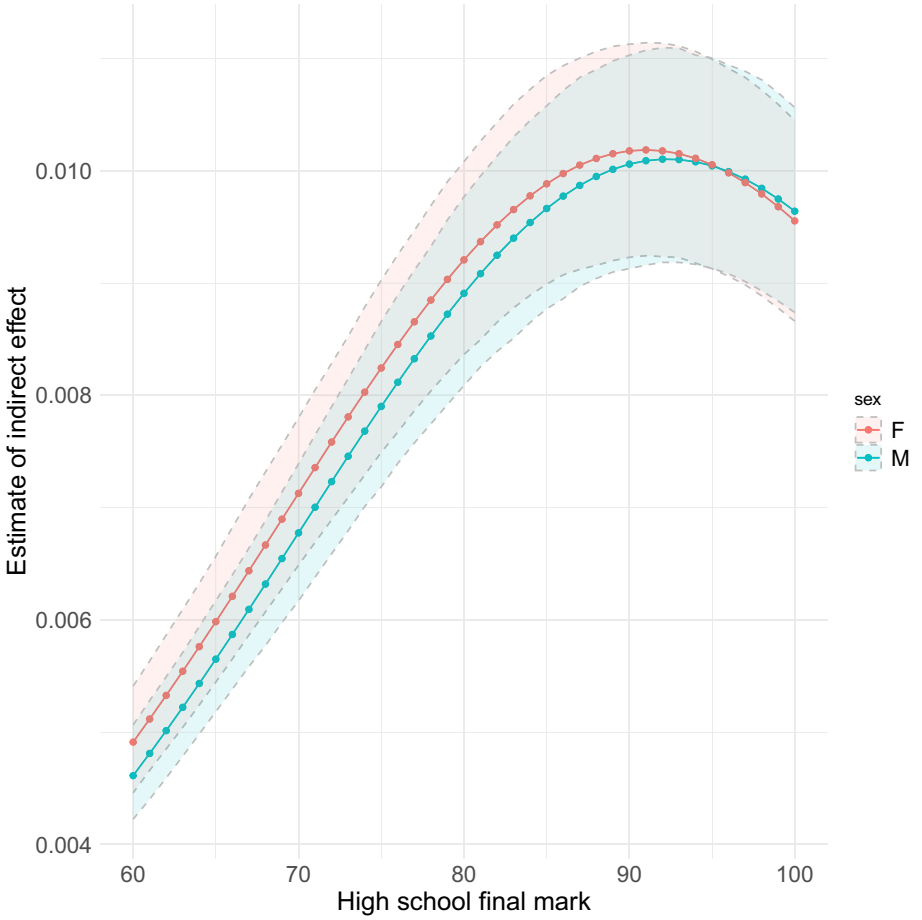

Fig. 5 CIEs for males and females and their confidence intervals estimated with the Montecarlo approach

Acknowledgements The authors are grateful to Massimo Attanasio, the Principal Investigator, and Andrea Priulla for granting access to the valuable data provided by the Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR), PRIN 2017 project titled 'From high school to job placement: micro data life course analysis of university student mobility and its impact on the Italian North–South divide' (Grant No. 2017HBTK5P).

Funding Open access funding provided by Università degli Studi di Palermo within the CRUI-CARE Agreement. The research work of Alessandro Albano has been partially supported by the European Union—NextGenerationEU—National Sustainable Mobility Center CN00000023, Italian Ministry of University and Research Decree n. 1033—17/06/2022, Spoke 2, CUP B73C2200076000.

Data Availability The data used in this study have been processed in accordance with the RESEARCH PROTOCOL FOR THE STUDY "From high school to the job placement: analysis of university careers and university mobility from Southern to Northern Italy" among the Ministry of University and Research, the Ministry of Education and Merit, the University of Palermo as the lead institution, and the INVALSI Institute. The reference researcher is Massimo Attanasio.

Declarations

Conflict of interest The authors report that there are no competing interests to declare.

Financial and non-financial interests The authors have no relevant financial or non-financial interests to disclose.

References

Albert, J.M.: Distribution-free mediation analysis for nonlinear models with confounding. Epidemiology **23**(6), 879–888 (2012)

Baron, R.M., Kenny, D.A.: The moderator–mediator variable distinction in social psychological research: conceptual, strategic and statistical considerations. J. Personal. Soc. Psychol. **51**(6), 1173–1182 (1986)

Bauer, D.J., Preacher, K.J., Gil, K.M.: Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: new procedures and recommendations. Psychol. Methods **11**(2), 142–163 (2006)

Berrar, D.P.: Performance measures for binary classification. In: Encyclopedia of Bioinformatics and Computational Biology (2019)

Biesanz, J.C., Falk, C.F., Savalei, V.: Assessing mediational models: testing and interval estimation for indirect effects. Multivar. Behav. Res. **45**(4), 661–701 (2010)

Bollen, K.A.: Structural Equations with Latent Variables. Wiley, New York (1989)

Curran, P.J.: Have multilevel models been structural equation models all along? Multivar. Behav. Res. **38**(4), 529–569 (2003)

Di Maria, C., Abbruzzo, A., Lovison, G.: Bayesian causal mediation analysis through linear mixed-effect models, Book of Short Papers-SIS 2022. Springer, Berlin (2022)

Doretti, M., Raggi, M., Stanghellini, E.: Exact parametric causal mediation analysis for a binary outcome with a binary mediator. Stat. Methods Appl. **31**, 87–108 (2022)

Efron, B., Tibshirani, R.J.: An Introduction to the Bootstrap. CRC Press, Boca Raton (1994)

Gaynor, S.M., Schwartz, J., Lin, X.: Mediation analysis for common binary outcomes. Stat. Med. **38**, 512–529 (2019)

Geldhof, G.J., Anthony, K.P., Selig, J.P., Mendez-Luck, C.A.: Accommodating binary and count variables in mediation: a case for conditional indirect effects. Int. J. Behav. Dev. **42**(2), 300–308 (2018)

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B.: Bayesian Data Analysis. CRC Press, Boca Raton (2013)

Hand, D.J., Till, R.J.: A simple generalisation of the area under the ROC curve for multiple class classification problems. Mach. Learn. **45**, 171–186 (2001)

Hayes, A.F., Preacher, K.J.: Quantifying and testing indirect effects in simple mediation models when the constituent paths are nonlinear. Multivar. Behav. Res. **45**, 627–660 (2010)

Kenny, D.A., Korchmaros, J.D., Bolger, N.: Lower level mediation in multilevel models. Psychol. Methods **8**(2), 115–128 (2003)

Koopman, J., Howe, M., Hollenbeck, J.R., Sin, H.P.: Small sample mediation testing: Misplaced confidence in bootstrapped confidence intervals. J. Appl. Psychol. **100**(1), 194–202 (2015)

Krull, J.L., MacKinnon, D.P.: Multilevel mediation modeling in group-based intervention studies. Eval. Rev. **23**(4), 418–444 (1999)

Krull, J.L., MacKinnon, D.P.: Multilevel modeling of individual and group level mediated effects. Multivar. Behav. Res. **36**(2), 249–277 (2001)

Loeys, T., Moerkerke, B., De Smet, O., Buysse, A., Steen, J., Vansteelandt, S.: Flexible mediation analysis in the presence of nonlinear relations: beyond the mediation formula. Multivar. Behav. Res. **48**(6), 871–894 (2013)
Lomnicki, Z.A.: On the distribution of products of random variables. J. R. Stat. Soc. Ser. B **29**(3), 513–524 (1967)
MacKinnon, D.P.: Introduction to Statistical Mediation Analysis. Taylor and Francis Group, New York (2008)
MacKinnon, D.P., Dwyer, J.H.: Estimating mediated effects in prevention studies. Eval. Rev. **17**, 144–158 (1993)
Miočević, M., MacKinnon, D.P., Levy, R.: Power in Bayesian mediation analysis for small sample research. Struct. Equ. Model. **24**(5), 666–683 (2017)
MOBYSU.IT. 2017. Database MOBYSU.IT, Mobilità degli studi universitari italiani, Research Protocol MUR—Universities of Cagliari, Palermo, Siena, Torino, Sassari, Firenze, Cattolica and Napoli Federico II, Scientific Coordinator Massimo Attanasio (UNIPA), Data Source ANS-MUR/CINECA
Morgan, S.L., Winship, C.: The Counterfactual Model. Analytical Methods for Social Research, pp. 31–58. Cambridge University Press, Cambridge (2007)
Pearl, J.: Direct and indirect effects. In: Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, UAI'01, San Francisco, CA, USA, pp. 411–420. Morgan Kaufmann Publishers Inc (IO) (2001)
Pearl, J.: Causal inference in statistics: an overview. Stat. Surv. **3**, 96–146 (2009a)
Pearl, J.: Causality. Cambridge University Press, Cambridge (2009b)
Pearl, J.: The causal mediation formula-a guide to the assessment of pathways and mechanisms. Prev. Sci. **13**(4), 426–436 (2012a)
Pearl, J.: The mediation formula: a guide to the assessment of causal pathways in nonlinear models, Chapter 12, pp. 151–179. Wiley, New York (2012b)
Preacher, K.J., Zyphur, M.J., Zhang, Z.: A general multilevel SEM framework for assessing multilevel mediation. Psychol. Methods **15**(3), 209–233 (2010)
Preacher, K.J., Zhang, Z., Zyphur, M.J.: Alternative methods for assessing mediation in multilevel data: the advantages of multilevel SEM. Struct. Equ. Model. **18**(2), 161–182 (2011)
Priulla, A.: Inequalities in student performances in the Italian universities. PhD thesis, University of Palermo (2023). Available at https://iris.unipa.it/handle/10447/582705
Rijnhart, J.J.M., Valente, M.J., MacKinnon, D.P., Twisk, J.W.R., Heymans, M.W.: The use of traditional and causal estimators for mediation models with a binary outcome and exposure-mediator interaction. Struct. Equ. Model. **28**(3), 345–355 (2021)
Rijnhart, J.J.M., Valente, M.J., Smyth, H.L., MacKinnon, D.P.: Statistical mediation analysis for models with a binary mediator and a binary outcome: the differences between causal and traditional mediation analysis. Prev. Sci. **24**(3), 408–418 (2023)
Rubin, D.B.: Causal inference using potential outcomes: design, modeling. Decis. J. Am. Stat. Assoc. **100**(469), 322–331 (2005)
Rubinstein, R.Y., Kroese, D.P.: Simulation and the Monte Carlo Method. Wiley, New York (2016)
Schluchter, M.D.: Flexible approaches to computing mediated effects in generalized linear models: generalized estimating equations and bootstrapping. Multivar. Behav. Res. **43**(2), 268–288 (2008)
Smithson, M., Verkuilen, J.: A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. Psychol. Methods **11**(1), 54 (2006)
Springer, M.D., Thompson, W.E.: The distribution of products of independent random variables. SIAM J. Appl. Math. **14**(3), 511–526 (1966)
Stolzenberg, R.M.: The measurement and decomposition of causal effects in nonlinear and nonadditive models. Sociolog. Methodol. **11**, 459–488 (1980)
Tsai, T.L., Shau, W., Hu, F.: Generalized path analysis and generalized simultaneous equations model for recursive systems with responses of mixed types. Struct. Equ. Model. **13**(2), 229–251 (2006)
Valeri, L., VanderWeele, T.J.: Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychol. Methods **18**(2), 137–150 (2013)
VanderWeele, T.: Explanation in Causal Inference: Methods for Mediation and Interaction. Oxford University Press, Oxford (2015a)
VanderWeele, T.J.: Explanation in Causal Inference. Oxford University Press, New York (2015b)
Wright, S.: The method of path coefficients. Ann. Math. Stat. **5**(3), 161–215 (1934)
Yuan, Y., MacKinnon, D.P.: Bayesian mediation analysis. Psychol. Methods **14**(4), 301–322 (2009)