Many Roads to Mediation: A Methodological and Empirical Comparison of Different Approaches to Statistical Mediation

Dominik Becker
Federal Institute for Vocational Education and Training (BIBB)

Abstract
This paper provides both a theoretical foundation and a simulation analysis of different statistical approaches to mediation. Regarding theory, a brief sketch of the fundamentals of mechanism-based explanations sets the argument of adhering to a consecutive order of predictor, mediator and outcome in mediation analysis. Having summarized the statistical fundamentals of different approaches to mediation analysis including simple mediation within OLS regressions, fixed-effects (FE) regressions, generalized-method-of-moments (GMM) regressions, causal mediation analysis without (CM) and with fixed effects (CMFE), and fixed-effects cross-lagged panel models (FE-CLPMs), I provide a simulation analysis with known but variable values for the intercorrelations between predictor, mediator and outcome in presence of unobserved heterogeneity and reverse causality. The aim of the simulation study is to examine differences in the relative performance of the aforementioned statistical approaches to mediation under different scenarios of causal order. Results reveal that OLS estimates are generally upwardly biased, FE and CMFE estimates by trend downwardly biased, and the ones of CM models (without FEs) can be biased in both directions. In contrast, coefficients and confidence intervals estimated by both GMM regressions and FE-CLPMs are most accurate – particularly if the structure of lags in the empirical models met the consecutive order set up in the data-generating process. Furthermore, FE-CLPMs are least sensitive to whether the first lag of the outcome variable is included as an additional predictor. All in all, analyses imply the importance that researchers most carefully translate their theoretical assumptions into an empirical model with the appropriate causal order.

Keywords: Panel data, Mediation, Unobserved heterogeneity, Reverse causality, Simulation analysis

© The Author(s) 2023. This is an Open Access article distributed under the terms of the Creative Commons Attribution 3.0 License. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Whether an observed association between two social constructs is based on a causal effect is one of the most fundamental methodological questions in the social sciences. Apart from simply asking if $X$ causes $Y$, social scientists are concerned with how a causal effect is brought about. From a theoretical perspective, this relates to the idea of a social mechanism $M$ (Hedström & Swedberg, 1996) along which an effect of $X$ on $Y$ is transmitted ($X \Rightarrow M \Rightarrow Y$). Statistically, this perspective translates into the broad field of mediation analysis which investigates whether a significant parameter estimate from some type of regression of $Y$ on $X$ persists once $M$ is controlled for. Also, it is possible to specify the share of the $X \Rightarrow Y$ effect that is transmitted via $M$ (“indirect” effect via the mediator), and the residual part (“direct effect”; Baron & Kenny, 1986).

When it comes to the identification of mediation effects in panel data, (at least) two important challenges need to be considered: First, if unobserved heterogeneity of either time-constant or time-varying covariates which are exogenous either to $X$ or to $M$ is present, the seeming mediation effect may be spurious (Imai et al., 2010). Second, a proper measurement of the causal order underlying the $X \Rightarrow M \Rightarrow Y$ chain must ensure that no reverse causality (in terms of current values of $X$ and/or $M$ being endogenous to prior values of $Y$) is present.

The aim of this paper is to explore how well different statistical approaches to mediation analysis are capable of addressing problems of causal order in the presence of unobserved heterogeneity with simulated data. In a brief theoretical section, I will first outline how the idea of mediation analysis relates to the social mechanisms approach to causality in the social sciences. I will then summarize different statistical approaches to mediation analysis and how they address problems of unobserved heterogeneity and reverse causality. Concretely, I will start with the simple “covariate inclusion” approach to mediation analysis in Ordinary Least Squares (OLS) regression. I will then move on to discuss how the introduction of (person) fixed-effects (FE) may solve problems of time-constant unobserved heterogeneity in panel data. A further extension, the Generalized Method of Moments (GMM), the most prominent of which is the Arrelano-Bond (AB) estimator (Arellano & Bond, 1991), additionally addresses the challenge of reverse causality by instrumenting both predictors and outcome by their respective lagged values of first, second, or higher order. A different approach to mediation is given by the causal mediation (CM) approach (Imai, Keele, Tingley, & Yamamoto, 2011) which advances Rubin’s (1986) potential outcomes (PO) model by the introduction of potential outcomes for the mediator variable giving treatment status on the one hand, and for the outcome given treatment and mediator status on the other.
hand. As this model has primarily been developed for cross-sectional data, it will prove useful to investigate its applicability to the analysis of panel data. Finally, I will discuss a more recent version of Fixed-Effects Cross-Lagged Panel Models (FE-CLPMs) which addresses both unobserved heterogeneity and reverse causality in the Structural Equation Modeling (SEM) framework (Allison, Williams, & Moral-Benito, 2017).

As the crucial touchstone of this study, I put all of the aforementioned approaches to mediation analysis to the test of an in-depth simulation analysis. Concretely, I will build on Leszczensky and Wolbring’s (2019) simulation study to generate random data with known but variable parameters for intercorrelations between $X$, $M$, and $Y$ in the presence of both unobserved heterogeneity and reverse causality. I will then explore how well different statistical approaches to mediation analysis can approximate the ‘true’ parameters. Finally, in the conclusion section, I will summarize the relative advantages of one analysis method over the other and provide practical recommendations in light of the theoretical idea of mediation.

## Theoretical Background

### Causality and Social Mechanisms

As statistical techniques matured over the course of the 20th century, it has been criticized that the quantitative approach might have gotten lost in “variable sociology”, i.e., a mainly data- and model-driven enterprise that lost sight of trying to ‘understand’ (e.g., Esser, 1996). Luckily, since the 1990s, mainly quantitative sociologists began to place renewed emphasis on the “understanding” dimension of explanation. One prominent proposition is grounded in the philosophy of social (but also life) science and posits a mechanism-based approach to explanation in the social sciences (Hedström, 2005; Hedström & Swedberg, 1996).

There exist numerous definitions of social mechanisms (Hedström & Ylikoski, 2010), the common denominator of which can be described as follows: “Social mechanisms are abstract and general models of spatially, temporally, and functionally organized entities and activities that explain why and how social phenomena are generated by preceding causal factors” (Tranow, Beckers, & Becker, 2016, 5f.; my emphasis).

Methodologically, the conceptual idea of a social mechanism as an explanation of why and how social phenomena are generated by preceding causal factors is closely related to the idea of statistical mediation. Consider the mechanism of “wishful thinking” (Elster, 1989): the desire for something to be true influences my belief about whether it is actually true and, in consequence, my correspond-
More generally, the impact of desires $D_i$ on action $A_i$ is brought about via (or, statistically speaking, mediated by) beliefs $B_i$ (Figure 1, Panel A). Continuing the above example, the effect of a better’s team preference on betting investments would be mediated by the subjective winning chances that the better attributes to their preferred team. But the mechanism approach is also suited to mapping the ideas of unobserved heterogeneity and reverse causality: With respect to unobserved heterogeneity, let $O_p$ refer to an unobserved component of the opportunity structure ($O$) (e.g., changes in shadow prices) which is prior ($p$) to both individuals’ desires $D_i$, beliefs $B_i$, and their corresponding action $A_i$. Let us further assume that $O_p$ brings about $D_i$, $B_i$, and $A_i$. In that case, we would not call desires $D_i$ a social mechanism with causal force (Figure 1, Panel B). Similarly, let us assume that $B_p$ refers to an (even observable) prior instance of belief $B_i$ which brings about desires $D_i$. In this case of reverse causality and in contrast to the general idea of wishful thinking (cf. panel A), $D_i$ would rather be a mechanism (or statistically: mediator) of $B_p$ effects on $A_i$ (Figure 1, Panel C).²

**Statistical Approaches to Mediation Analysis**

**Simple mediation**

A seminal definition of mediation analysis was formulated by Baron and Kenny (Baron & Kenny, 1986, p. 1177; also see Figure 2):

---

1 For the DBO scheme linking individuals’ desires and beliefs to situational opportunities see Hedström (2005).

2 There exist of course other forms of heterogeneity that might complicate the identification of mediation effects. Below, I will only briefly touch upon these issues as they surpass what will be covered in the simulation analyses presented below, but I will advise directions for future research in the conclusion section.
“A variable functions as a mediator when it meets the following conditions: (a) variations in levels of the independent variable significantly account for variations in the presumed mediator (i.e., Path a), (b) variations in the mediator significantly account for variations in the dependent variable (i.e., Path b), and (c) when Paths a and b are controlled, a previously significant relation between the independent and dependent variables is no longer significant, with the strongest demonstration of mediation occurring when Path c is zero.”

It is further common to distinguish between a direct, an indirect, and the total effect of a predictor (or treatment) variable on its outcome. In Figure 2, the direct effect is given by path c, the indirect effect is the product of paths a and b, and the total effect is the sum of both the direct and the indirect effect, i.e. \( c + a \times b \) (Hayes, Preacher, & Myers, 2011, p. 438).

Consequently, a rigorous application of the simple mediation model in regression analysis would first estimate the effect of an independent variable \( X \) on the potential mediator variable \( M \) to ensure that Baron and Kenny’s (1986) condition a) is met:

\[
M_{(i)} = \beta_{0M} + \beta_1X_{i} + \epsilon_{M(i)}.\tag{1}
\]

In a second step, the dependent variable of interest \( Y \) is predicted by \( X \) (2), and in a third step, by both \( X \) and \( M \) (3) to explore whether the effect of \( X \) on \( Y \) persists once (2) is controlled for \( M \). In practice, both (2) and (3) will often add a vector of covariates \( C \) to ensure that neither the relation of \( X \) nor the one of \( M \) to \( Y \) is spurious:

\[
Y_{i} = \beta_{0Y} + \beta_2X_{i} + \beta_3C_{i} + \epsilon_{Y(i)},\tag{2}
\]

\[
Y_{i} = \beta_{0Y} + \beta_2X_{i} + \beta_3C_{i} + \beta_4M_{i} + \epsilon_{Y(i)}.\tag{3}
\]

Both unobserved heterogeneity and reverse causality can be addressed in the simple mediation model once we assume to have panel data at our disposal. In that case, unobserved heterogeneity can be addressed using (person-level) fixed
effects (FEs) which ‘de-mean’ both $X$ and $Y$ to remove any variation between individuals which is constant over time (e.g., gender, migration background, or the fixed part of personality differences). Adding subscript $t$ to refer to observation time, equation (3) amounts to

$$Y_{i(t)} - \bar{Y}_t = \beta_{0Y} + \beta_2(X_{i(t)} - X_t) + \beta_3(C_{i(t)} - \bar{C}_t) + \beta_4(M_{i(t)} - \bar{M}_t)$$

$$+ (\alpha_t - \bar{\alpha}_t) + \epsilon_{Yi(t)} - \bar{\epsilon}_{Yi(t)}.$$  

(4)

Since $\alpha_t$ is time-constant by definition, it is identical to its person-specific mean. Consequently, $(\alpha_t - \bar{\alpha}_t)$ amounts to zero, and unobserved heterogeneity is wiped out after demeaning.

FE regressions build on the assumption of strict exogeneity, meaning that current values of $\epsilon_{Yi(t)}$ should not depend on past, present and future values of $X_{it}$ (Brüderl & Ludwig, 2015). This assumption is violated in the case of reverse causality, i.e., when $Y_{i(t)}$ affects $X_{i(t+1)}$ (Leszczensky & Wolbring, 2019). As a consequence, estimates of (4) will be biased if reverse causality is present. To address this issue, researchers often apply ‘lags’ to $X$ or $M$, i.e., they use observations one or even more waves prior to the one in which $Y$ is observed. In accordance to the idea of a causal order in terms of changes in $X$ affecting changes in $Y$ via changes in $M$, one approach could be to predict $Y_{it}$ via $X_{i(t-2)}$ and $M_{i(t-1)}$, i.e., applying the first lag to the mediator of interest, and the second lag to the main predictor at hand:

$$Y_{i(t)} - \bar{Y}_t = \beta_{0Y} + \beta_2(X_{i(t-2)} - X_t) + \beta_3(C_{i(t)} - \bar{C}_t) + \beta_4(M_{i(t-1)} - \bar{M}_t)$$

$$+ \epsilon_{Yi(t)} - \bar{\epsilon}_{Yi(t)}.$$  

(5)

However, it has been shown both analytically and based on simulations that lags of either variable do not circumvent biased estimates and statistical inference in the case of reverse causality (Reed, 2015). A more generalized approach that also relies on lagged variables, but tries to resolve identification issues of previous approaches, is the Generalized Method of Moments (GMM), a particular version of which is known as the Arellano-Bond (AB) estimator (Arellano & Bond,

---

3 There are several methods to address the problem of unobserved heterogeneity in panel data: *first-differences*, where each current value of a variable is subtracted by the one of the previous wave, *person dummies*, which include dummy variables for all $n-1$ individuals in the sample, and *demeaning*, where each value of a variable is subtracted by its unit-specific mean over time. The latter approach is explained more extensively below and is also the one that will be used in the simulation study to follow.
In its most simplistic form, the AB approach starts from the following model:\(^4\)

\[ Y_{it} = \beta_1 Y_{i(t-1)} + \beta_2 X_{it} + \alpha_i + \epsilon_{i(t)}. \]  

(6)

As a first step, first-differences for all terms in (6) are computed to get rid of time-
constant unobserved heterogeneity \( \alpha_i \):

\[ \Delta Y_{it} = \beta_1 \Delta Y_{i(t-1)} + \beta_2 \Delta X_{it} + \Delta \epsilon_{i(t)}. \]  

(7)

As a second step, \( Y_{i(t-2)} \) is used as an instrument for \( \Delta Y_{i(t-1)} \). In practice, and as
recommended by the authors, additional higher-order lags of \( Y \) (\( \Delta Y_{i(t-3)}, \Delta Y_{i(t-4)}, \)...
are often used to instrument \( \Delta Y_{i(t-1)} \) (Arellano & Bond, 1991). Alternatively,
or in addition, \( \Delta Y_{i(t-1)} \) may be instrumented by second, third, or even higher-order
differences of \( Y \) (\( \Delta Y_{i(t-2)}, \Delta Y_{i(t-3)}, \) ...). By this design, it is possible to separate
strictly exogenous from sequentially exogenous or predetermined variables from
one another. Consequently, “AB-type panel estimators thus weaken the exogene-
ity assumption for a subset of regressors, thereby providing consistent estimates
even if reverse causality is present” (Leszczensky & Wolbring, 2019, p. 9).

Yet, despite this pleasant statistical property, real-world applications of the
AB estimator are not without pitfalls: As Allison et al. (2017) outline, while the
AB-estimator provides consistent estimators, “there is evidence that the estimators
are not fully efficient, have considerable small-sample bias, and often perform
poorly when the autoregressive parameter (the effect of a variable on itself
at a later point in time) is near 1.0” (p. 1f.). In my discussion of the FE-CLPM,
I will come back to how these drawbacks may be circumvented by a maximum-
likelihood approach.

**Causal mediation analysis**

Imai, Keele, et al. (2011) advance the idea of mediation analysis as a method-
ological approximation to causal mechanisms within the potential outcomes (PO)
framework (Rubin, 1986). In contrast to previous common practice when social
scientists tended to interpret each estimate of multivariate analysis as causal, the
PO approach focuses on the causal identification of solely one effect, called treat-
ment \( T \), on the outcome of interest, \( Y \). Although the question of how a particular
individual \( i \) in the treatment group would have behaved had they not received
the treatment cannot be answered empirically, it can be approximated by com-
paring outcome \( Y \) of the treatment group (\( Y_{i|T=1} \)) with the non-treatment group
(\( Y_{i|T=0} \)):

---

\(^4\) As a distinct AB-type equation for the mediator is not shown, subscript \( Y \) is omitted for
now.
\[ T_i \equiv Y_i(1) - Y_i(0). \] (8)

The next step is to introduce the mediator variable into the PO main equation. For dichotomous mediators, outcome \( Y \) in the treatment group under the condition of \( M=1 \) (\( Y_i | T=1, M=1 \)) is compared to \( Y \) in the non-treatment group under the condition of \( M=0 \) (\( Y_i | T=0, M=0 \)):

\[ T_i \equiv Y_i(1, M_i(1)) - Y_i(0, M_i(0)). \] (9)

Having defined mediation in the PO framework, it is possible to define the indirect or causal mediation effect

\[ \delta_i(t) \equiv Y_i(t, M_i(1)) - Y_i(t, M_i(0)). \] (10)

which refers to paths a) and b) in Figure 2, as well as the direct/residual effect

\[ \zeta_i(t) \equiv Y_i(1, M_i(t)) - Y_i(0, M_i(t)) \] (11)

which amounts to path c) in Figure 2.

Another important assumption for causal mediation in the potential outcomes framework is the one of sequential ignorability (SIA), which can be decomposed into ignorability of treatment assignment (ITA) given \( X \),

\[ Y_i(t', m), M_i(t) \perp T_i \mid X_i = x, \] (12)

and ignorability of mediator status (IMS) given \( T + X \):

\[ Y_i(T, m) \perp M_i(t) \mid T_i = t, X_i = x. \] (13)

Concretely, ITA given \( X \) in (12) means that having controlled for a vector of covariates (which is here denoted \( X \)), it should be random whether a particular individual \( i \) belongs to the treatment or to the control group. Furthermore, IMS given \( T \) and \( X \) in (13) means once I know whether individual \( i \) belongs to the treatment or to the control group and I have controlled for my set of covariates \( X \), there should (by assumption) be no other systematic variation in the mediator variable.

How are unobserved heterogeneity and reverse causality addressed in the causal mediation model? Regarding unobserved heterogeneity, the SIA is crucial: If the set of covariates \( C \) is exhaustive and both treatment and mediator status are independent of unmeasured confounders, unobserved heterogeneity is no issue by definition. For particular scenarios in which the causal effect of \( T \) on \( Y \) is passed on across a second, unobserved mediator \( N \) that either runs parallel to the observed mediator \( M \) or is endogenous to the latter (Figure 3, Panel A; taken by Imai, Keele, et al., 2011, p. 786), the SIA is violated but can yet be addressed via sensitivity analyses in which the correlation between the residual terms of
both the mediator and the outcome equation is examined (Imai, Keele, & Tingley, 2010; Imai, Keele, & Yamamoto, 2010). For that purpose, it is useful to specify mediation in the linear structural equation framework again (Imai, Keele, & Yamamoto, 2010, p. 57; Imai, Keele, et al., 2011, p. 774): In our notation (cf. equations (1) and (2)), the correlation of interest is defined as $\rho = \text{corr}(\epsilon_Y(i), \epsilon_M(i))$. The magnitude of $\rho$ can be used to measure to what extent the SIA is violated: in the case of no violation, $\rho$ should amount to zero; the more severely the model deviates from this ideal state, the larger $\rho$. The key element of the sensitivity analysis is now to approximate the unobserved mediator by a random variable whose correlations with $T$, $M$ and $Y$ are varied over the course of the estimation process. As an alternative measure of potential bias due to an unobserved mediator, relative changes in $R^2$ can be used. In contrast, the case of $M$ being endogenous to an unobserved mediator $N$ constitutes a severe threat to the SIA and cannot be addressed by sensitivity analyses (Figure 3, Panel B).

Concerning reverse causality between $T$, $M$ and $Y$, the causal mediation proponents simply state that “[l]ongitudinal data with covariates (realized and measured before treatment assignment) and treatment assignment (realized and measured before outcomes) eliminates the possibility of reverse causality and thus provides a clear way to adhere to this prescription of design followed by analysis” (Imai, Jo, & Stuart, 2011, p. 868). Since it is well known, however, that a discrete longitudinal measurement of relevant indicators (i.e., in terms of annual panel waves) is no insurance against unobserved forms of reverse causality (Leszczensky & Wolbring, 2019), it remains an open question as to how the causal mediation approach can handle this challenge. I will address this issue in my simulation analysis section.5

---

5 Lutz, Sordillo, Hokanson, Chen Wu, and Lange (2020) provide a first insight into how sensitively the causal mediation approach reacts to reverse causality. However, they do not consider the case in which both unobserved heterogeneity and reverse causality is present simultaneously.
SEM approach to mediation

The SEM approach to mediation advances the simple mediation model both structurally and in terms of measurement: First, as longitudinal data is structurally arranged in ‘wide’ format, more complex meditational structures (e.g., two mediators at once) can be easily implemented. Second, the SEM approach holds a more elaborate perspective on the measurement component of the constructs at hand, which amounts to the option of using latent variable models for both predictor variable(s), mediator(s), outcome(s), and covariates. As for the ease of comparison between mediation approaches I will refrain from using latent variable models in the simulation models; the formal details to follow will focus on observed variable models which are just a special case of latent variable models.

For a conventional “x ’causes’ y” model without any mediator, the structural part is defined as in conventional OLS regression analysis (cf. Bollen, 1989, 41ff.):

$$ Y = \gamma_1 X + \zeta_1, $$

where $Y$ denotes the dependent variable, $X$ the independent variable with regression weight $\gamma_1$ on $Y$, and $\zeta_1$ the error, residual or disturbance term.

As before, a mediator variable $M$ can be introduced by setting it exogenous to $Y$ and endogenous to $X$:

$$ M = \gamma_2 X + \zeta_2, $$

$$ Y = \gamma_1 X + \gamma_3 M + \zeta_1. $$

As usual, the indirect effect for observed variable models is defined as the difference between the total effect and the direct effects. For latent variable models, the decomposition of direct, indirect and total effects is more complex (see Bollen, 1989, 376ff.). Luckily, modern statistical software which is capable of estimating SEMs – such as Stata, R (with lavaan in particular) or Mplus – provides handsome sub-routines to decompose total, direct and indirect effects in both observed and latent variable models (see, e.g., Mehmetoglu, 2018; Muthén, 2017; Rosseel, 2012).

While the added value of mediation of observed variables within the SEM approach may not be evident at first sight, its advantage becomes more obvious when it comes to addressing the challenge of reverse causality in panel data. There is a long tradition within the SEM approach to do so by means of cross-lagged panel models (CLPMs; also see Finkel, 1995). Taking advantage of the wide data structure underlying the SEM approach, in case of a predictor $X$ and an outcome $Y$ measured at times $t_1$ and $t_2$, a cross-lagged panel model applies the following steps:
That is, $Y_2$ is regressed on both $X_1$ and $Y_1$, while at the same time, $X_2$ is regressed on both $X_1$ and $Y_1$. Apart from simply controlling for potential reverse causality effects, one appeal of the CLPM is that reciprocal effects which are often assumed by theory can be directly estimated (Selig & Little, 2012, p. 268). A crucial objection that has been raised against the CLPM is that it may lead to biased results in case of unobserved stable individual-level characteristics (Hamaker, Kuiper, & Grasman, 2015). There have already been several approaches to incorporate the FE estimator into the SEM framework both with and without a cross-lagged structure (Allison, 2009; Curran & Bollen, 2001). A more recent approach to Fixed-Effects Cross-Lagged Panel Models (FE-CLPMs) by Allison et al. (2017) draws on previous work of Moral-Benito (2013) who has outlined a maximum-likelihood-based estimation method that circumvents several computational drawbacks of GMM estimators in general and of the AB method in particular. The contribution of Allison et al. (2017) is to integrate Moral-Benito’s (2013) approach into the general SEM framework, as a consequence of which it can be estimated using conventional SEM software subroutines.

The FE-CLPM is defined as follows:

\[
Y_{i(t)} = \mu(t) + \beta_1 X_{i(t\_t-1)} + \beta_2 Y_{i(t\_t-1)} + \delta_1 W_{i(t)} + \gamma_1 Z_i + \alpha_1 + \epsilon_{i(t)},
\]

\[
X_{i(t)} = \tau(t) + \beta_3 X_{i(t\_t-1)} + \beta_4 Y_{i(t\_t-1)} + \delta_2 W_{i(t)} + \gamma_2 Z_i + \eta_1 + \nu_{i(t)}.
\]

where in (19) $\mu_t$ describes the intercept of $Y$ that varies across time $t$, $\beta_1$ and $\beta_2$ are scalar coefficients assessing how $Y$ is predicted by former values of both $X$ and $Y$, $\delta_1$ and $\gamma_1$ are row vectors of coefficients for both time-variant controls variables $W$ and time-constant control variables $Z$, $\alpha_1$ refers to the joint effects of time-constant unobservables (assuming them to exert constant effects on $Y_{i(t)}$, and $\epsilon_{i(t)}$ is a random error term.

Accordingly, in (20), $\tau_{(t)}$ describes the intercept of $X$ that varies across time $t$, $\beta_3$ and $\beta_4$ are scalar coefficients assessing how $X$ is predicted by former values of both $X$ and $Y$, $\delta_2$ and $\gamma_2$ are row vectors of coefficients for both time-variant controls variables $W$ and time-constant control variables $Z$, $\eta_1$ refers to the joint effects of time-constant unobservables (assuming them to exert constant effects on $X_{i(t)}$, and $\nu_{i(t)}$ is a random error term.

The most notable difference compared to the ‘traditional’ CLPM presented in (17)-(18) is the inclusion of terms $\alpha_1$ and $\eta_1$ to address time-constant unobserved effects on $Y_{i(t)}$ and $X_{i(t)}$, respectively. In econometric approaches, $\alpha_1$ and $\eta_1$ are often assumed to be “fixed”, i.e., exert the same effect for each individual, whereas in other social science disciplines, this assumption might be relaxed (e.g., Hamaker et al., 2015).
To recall, a combination of fixed effects and lagged outcome variables will lead to biased estimates of the $\beta$ coefficients. Within the AB approach, this has been addressed by, first, removing fixed effects by computing first differences for $X$ and $Y$, and then, second, instrumenting these differences by lagged difference scores (cf. eq. (7)), which are finally, third, estimated by GMM. It is well-known, however, that GMM approaches are particularly sensitive to the number of lags and corresponding instruments (Leszczensky & Wolbring, 2019; Roodman, 2009). In contrast, the ML approach to reverse causality produces estimators that are asymptotically equivalent to GMM, but have more preferable finite sample properties in case of weak and/or numerous instruments (Moral-Benito, 2013).

In what follows, Allison et al. (2017) argue that the ML approach to the cross-lagged model with fixed effects is a special case of the general SEM framework outlined in (12) which is illustrated in Figure 4. Leaving aside both $W$ and $Z$ variables and focusing on the case of manifest $X$ and $Y$ the latter of which is measured on four occasions, it is evident that while $Y_t$ is predicted by $Y_{t-1}$, this is not the case for instances of $X$ which are simply allowed to correlate with one another. In addition, each $Y_t$ is predicted by $X_{t-1}$ as well as $\alpha_1$, which is the FE estimate intended to address time-constant unobserved heterogeneity. Coefficient $\alpha_1$, in turn, correlates with all instances of $X$ (but is not allowed to correlate with

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.png}
\caption{The FE-CLPM. Source: Allison et al. (2017, 6).}
\end{figure}
any time-invariant observable $Z$ if the latter is present in the model.\footnote{As a consequence of this identificatory step, it is advised to exclude all time-constant variables from the estimation model (Allison et al. 2017: 6).} Finally, and of crucial importance, $x_3$ is allowed to correlate with $\epsilon_2$, the error term of $Y_2$. According to Allison et al. (2017, 6), it is this correlation that makes $X$ predeter-
mined (by $Y$). In other words, this correlation is the crucial leverage to account for reverse causality between $X$ and $Y$.

**Observed heterogeneity and interim conclusion**

Apart from the challenges of reverse causality and unobserved heterogeneity, the statistical approaches just discussed can also address several issues of observed heterogeneity. There are different terms by which this kind of heterogeneity is referred to, the most prominent of which are interaction effects, moderator effects, multiplicative effects, and treatment effect heterogeneity (Baron & Kenny, 1986; Brambor, Clark, & Golder, 2006; Xie, Brand, & Jann, 2012). As a common denominator, a predictor (or treatment) variable is multiplied (i.e., “interacted”), with an observed variable $Z$. In our case, we can generally distinguish three possible interaction terms: $i$) between the main predictor (or treatment) variable (usually denoted $X$ or $T$) and another moderating variable $Z$; $ii$) between the mediator $M$ and $Z$, and between $X$ (or $T$) and $M$. It can be formally outlined that the above approaches are generally capable to address either form of observed heterogeneity (available upon request). In contrast, and as outlined above, they differ in their capacity to address unobserved heterogeneity and reverse causality. The essence of this methodological comparison is tabulated in Table 1.
Table 1  Comparison of different statistical approaches to mediation analysis in their capacity to address several methodological challenges

<table>
<thead>
<tr>
<th></th>
<th>Observed heterogeneity</th>
<th>Unobserved heterogeneity</th>
<th>Reverse causality</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OLS</strong></td>
<td>Can incorporate interac-</td>
<td>Not in baseline model,</td>
<td>May incorporate lags of</td>
</tr>
<tr>
<td></td>
<td>tions of type $XZ$, $MZ$, and $XM$</td>
<td>but can be advanced to</td>
<td>$X$ and $Y$, but results will</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FE estimator by manual</td>
<td>be biased</td>
</tr>
<tr>
<td></td>
<td></td>
<td>demeaning</td>
<td></td>
</tr>
<tr>
<td><strong>FE</strong></td>
<td>Can incorporate interac-</td>
<td>Rules out time-constant</td>
<td>May incorporate lags of</td>
</tr>
<tr>
<td></td>
<td>tions of type $XZ$, $MZ$, and $XM$</td>
<td>unobserved heterogeneity</td>
<td>$X$ and $Y$, but results will</td>
</tr>
<tr>
<td></td>
<td></td>
<td>by demeaning all variables</td>
<td>be biased</td>
</tr>
<tr>
<td><strong>AB/GMM</strong></td>
<td>Can incorporate interac-</td>
<td>See FE</td>
<td>First-differences for $X$</td>
</tr>
<tr>
<td></td>
<td>tions of type $XZ$, $MZ$, and $XM$</td>
<td></td>
<td>and $Y$ instrumented by</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>higher-order lags</td>
</tr>
<tr>
<td><strong>CM</strong></td>
<td>Can incorporate interac-</td>
<td>See OLS. Yet, empirical</td>
<td>May incorporate lags of</td>
</tr>
<tr>
<td></td>
<td>tions of type $XZ$ and $XM$ (unclear if $MZ$ identified)</td>
<td>performance of manual approach untested hitherto.</td>
<td>$X$ and $Y$, but empirical</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>performance of this approach untested hitherto.</td>
</tr>
<tr>
<td><strong>SEM</strong></td>
<td>Can incorporate interac-</td>
<td>Not in baseline model</td>
<td>Addressed by cross-lagged panel-model</td>
</tr>
<tr>
<td></td>
<td>tions of type $XZ$, $MZ$, and $XM$</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>FE-CLPM</strong></td>
<td>Can incorporate interac-</td>
<td>Introduces variables $\alpha$ and $\eta$ to capture</td>
<td>See SEM</td>
</tr>
<tr>
<td></td>
<td>tions of type $XZ$, $MZ$, and $XM$</td>
<td>unobserved heterogeneity effects on $X$ and $Y$, respectively</td>
<td></td>
</tr>
</tbody>
</table>

Simulation Analysis

The Present Study

Previous simulation studies have revealed that both OLS and FE analysis are biased when both unobserved heterogeneity and reverse causality are present (Leszczensky & Wolbring, 2019). Other research based on simulation analysis suggests that GMM strategies such as the AB estimator can run into problems, for instance, when the number of waves is small and lags are long (Newey & Windmeijer, 2009; Windmeijer, 2005). Further simulation studies suggest that the FE-CLPM can keep up with the GMM approach in the presence of both unobserved heterogeneity and reverse causality (Allison et al., 2017; Moral-Benito, Allison,
& Williams, 2019; also see Leszczensky & Wolbring, 2019). Yet, two gaps in research can be identified which the present contribution aims to address.

First, it has not yet been explored if these results generalize to the inclusion of a mediator variable which, in an ideal-world data-generating process (DGP), will be preceded by the main predictor but succeeded by the outcome (see below). Second, it has not been tested how the gold standard in mediation analysis, the causal mediation model in the potential-outcomes framework, performs if the challenges of unobserved heterogeneity and reverse causality are addressed by “on-board resources” in terms of demeaning and lagging all relevant variables.

Consequently, I will now present a simulation analysis to evaluate which of the statistical approaches to mediation analysis identifies the parameter values of predictor $X$, a mediator $M$, and their corresponding lags – which have been specified in the DGP prior to the simulation analysis – with minimal bias.

**Parameters and scenarios of the simulation model**

My simulation analysis builds on the one by Leszczensky and Wolbring (2019) but advances it by including an additional variable $M$ which shall mediate the effect of $X$ on $Y$ in the simulated data set. I first generated data with intercorrelations of $\rho_{(X,M,Y)} = .5$ and standard normally distributed independent error terms at $t_0$, respectively. This data was expanded to waves 1-5 in a second step by the following data-generating process (DGP):

\[
Y_{it} = \beta_1 Y_{it-1} + \beta_2 X_{it-2} + \beta_3 M_{it-1} + \beta_4 Z_i + \epsilon_{it} \quad \text{with} \quad \epsilon_{it} \sim N(0; 1),
\]

\[
X_{it} = \beta_5 Y_{it-1} + \beta_6 Z_i + \mu_{it} \quad \text{with} \quad \mu_{it} \sim N(0; 1),
\]

\[
M_{it} = \beta_7 Y_{it-1} + \beta_8 X_{it-1} + \beta_9 Z_i + \nu_{it} \quad \text{with} \quad \nu_{it} \sim N(0; 1).
\]

Above, $\beta_1$ refers to the extent of autocorrelation for outcome $Y$. As the variation of $\beta_1$ had no substantial impact on the simulation results by Leszczensky and Wolbring (2019), I set the parameter to be constant ($\beta_1 = .5$). Most important, $Y_{it}$ is modeled as an outcome of both $X_{it-2}$ (with effect $\beta_2$) and $M_{it-1}$ (with effect $\beta_3$). That is, in accordance to the idea of a social mechanism which is by definition situated between a cause and its outcome, the DGP understands the mediation model as the statistical pendent of a mechanism-based explanation. Consequently, the consecutive order of $X$, $M$ and $Y$ is of vital importance here. While Leszczensky and Wolbring (2019) switch between contemporaneous and lagged effects of $X$ on $Y$, my model is more simplistic in assuming constant effects of $X_{it-2}$ on $Y_{it}$.

In addition, $Z$ denotes an unmeasured, time-constant normally-distributed variable that addresses the challenge of unobserved heterogeneity. $Z$ is associated with $Y$, $X$, and $M$ by parameters $\beta_4$, $\beta_6$ and $\beta_9$, respectively. To simplify the
simulation model, these were set to 0.5 (unobserved heterogeneity moderately present), respectively. For all possible combinations of parameters (which are summarized in Table 2), 500 datasets with 500 observations each were generated.

**Models**

To compare point estimates and corresponding confidence intervals of the aforementioned mediation approaches, for either of them, the same set of sub-models will be estimated. Concretely, for both 1) FE regressions, 2) the GMM approach, 3) the causal mediation (CM) approach, and 4) the FE-CLPM, the following scenarios will be compared (see Table 3): *Scenario A*) employs a simultaneous analysis of $Y$ predicted by the variables at the same point in time. *Scenario B*) takes the first lag of all variables to predict later instances of $Y$. *Scenario C*) follows the idea of a consecutive order between $X$, $M$, and $Y$ (which is inspired by the rationale of mechanism-based explanations) by modeling $Y$ by the second lag of $X$ and the first lag of $M$. Finally, *scenario D*) amends *scenario C*) by adding the first lag of $Y$ to account for potential reverse causality between $X$ and $Y$.

Moreover, for each scenario, the following two submodels are estimated: *Submodel i*) predicts $Y$ only by $X$ (or its first or second lag) or, as in *scenario D*), the first lag of $Y$, and *submodel ii*) adds the mediator variable $M$ (or its first lag).

### Table 2 Parameter values of the simulation analysis

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Concept</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>Autocorrelation of $Y$</td>
<td>0.5</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>Effect of $X_{t-2}$ on $Y_t$</td>
<td>0, 0.5</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>Effect of $M_{t-1}$ on $Y_t$</td>
<td>0, 0.5</td>
</tr>
<tr>
<td>$\beta_8$</td>
<td>Effect of $X_{t-1}$ on $M_t$</td>
<td>0, 0.5</td>
</tr>
<tr>
<td>$\beta_4$, $\beta_6$, $\beta_9$</td>
<td>Unobserved heterogeneity on $Y$, $X$, $M$, respectively</td>
<td>0.5</td>
</tr>
<tr>
<td>$\beta_5$, $\beta_7$</td>
<td>Reverse causality on $X$ and $M$, respectively</td>
<td>0.5</td>
</tr>
</tbody>
</table>
Table 3  Scenarios for the simulation study

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Submodel i)</th>
<th>Submodel ii)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A) Simultaneous scenario (no lags)</td>
<td>$Y_t = X_t$</td>
<td>$Y_t = X_t + M_t$</td>
</tr>
<tr>
<td>B) Lagged scenario</td>
<td>$Y_t = X_{t-1}$</td>
<td>$Y_t = X_{t-1} + M_{t-1}$</td>
</tr>
<tr>
<td>C) Consecutive scenario</td>
<td>$Y_t = X_{t-2}$</td>
<td>$Y_t = X_{t-2} + M_{t-1}$</td>
</tr>
<tr>
<td>D) Consecutive scenario + LI(Y)</td>
<td>$Y_t = Y_{t-1} + X_{t-2}$</td>
<td>$Y_t = Y_{t-1} + X_{t-2} + M_{t-1}$</td>
</tr>
</tbody>
</table>

Results

Tables 4-6 show the results of the simulation study. Table 4 lists the predicted $\beta$ coefficients and their corresponding standard errors for both OLS and FE analyses of the simulated data. Between columns, it is differentiated between the four data simulation scenarios (see Table 3). Between rows, the values for the regression parameters are varied (see Table 2), and it is differentiated between two submodels one of which predicts $Y$ only by $X$, and the other one by both $X$ and $M$. If the predicted $\beta$ coefficients of $X$ and/or $M$ are subject to a bias of $|\varepsilon_\beta| > 0.1$, the background color of the corresponding table cell is highlighted in different shades of green for upward bias, and in different shades of red for downward bias (see the explanatory notes below Tables 4-6). In addition, Figures A1-A6 in Appendix A show coefficient plots of all parameter estimates and corresponding confidence intervals. These plots may provide visual aid to answer the question of if the statistical approaches applied to the simulation models correctly identify mediation effects which may or may not have been set in the underlying DGP.

For the OLS approach, when all $\beta$ coefficients have been set to zero, the predicted effects of $X$ and $M$ on $Y$ are overestimated given they have been set to be absent in the DGP (see left panel of Table 4). The upward bias within this particular setting is largest in the lagged scenario, and smallest in the consecutive scenario controlled for the first lag of $Y$. Once $\beta_2$ and/or $\beta_3$ are set to .5, this pattern persists for most of the predicted effects of $X$, and their bias is generally larger as long as the analyses have not controlled for $M$. If they do, the OLS approach incorrectly identifies mediation effects of $M$ although $\beta_8$ is still set to zero (see Appendix A, Figure A1a). Furthermore, if $\beta_8$ is set to .5, the amount of mediation predicted by the OLS approach is way too high particularly in case of $\beta_3 = .5$ (Appendix A, Figure A1b). The general upward bias of the OLS approach is most pronounced if both $\beta_2$ and $\beta_3$ are set to .5. In contrast, when both $\beta_2$ and $\beta_8$ are set to .5, predicted effects of $X$ may be slightly downwardly biased in the contemporary and lagged scenarios given they have not been controlled for $M$. 
Table 4  Results of Ordinary Least Squares (OLS) and Fixed-Effects (FE) regressions of simulated data

<table>
<thead>
<tr>
<th></th>
<th>OLS Regressions</th>
<th>FE Regressions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>contemporary</td>
<td>logged</td>
</tr>
<tr>
<td>( \beta_2 = \beta_3 = \beta_8 = 0 )</td>
<td>X (β) 0.375, M (β) 0.385</td>
<td>X (β) 0.375, M (β) 0.381</td>
</tr>
<tr>
<td>( \beta_2 = 0.5, \beta_3 = 0, \beta_8 = 0 )</td>
<td>X (β) 0.252, M (β) 0.263</td>
<td>X (β) 0.256, M (β) 0.217</td>
</tr>
<tr>
<td>( \beta_2 = 0, \beta_3 = 0.5, \beta_8 = 0 )</td>
<td>X (β) 0.294, M (β) 0.317</td>
<td>X (β) 0.244, M (β) 0.169</td>
</tr>
<tr>
<td>( \beta_2 = 0.5, \beta_3 = 0.5, \beta_8 = 0 )</td>
<td>X (β) 0.292, M (β) 0.301</td>
<td>X (β) 0.236, M (β) 0.140</td>
</tr>
<tr>
<td>( \beta_2 = 0, \beta_3 = 0, \beta_8 = 0.5 )</td>
<td>X (β) 0.552, M (β) 0.543</td>
<td>X (β) 0.674, M (β) 0.545</td>
</tr>
<tr>
<td>( \beta_2 = 0.5, \beta_3 = 0, \beta_8 = 0.5 )</td>
<td>X (β) 0.502, M (β) 0.503</td>
<td>X (β) 0.700, M (β) 0.576</td>
</tr>
<tr>
<td>( \beta_2 = 0, \beta_3 = 0.5, \beta_8 = 0.5 )</td>
<td>X (β) 0.170, M (β) 0.187</td>
<td>X (β) 0.134, M (β) 0.093</td>
</tr>
<tr>
<td>( \beta_2 = 0.5, \beta_3 = 0.5, \beta_8 = 0.5 )</td>
<td>X (β) 0.237, M (β) 0.230</td>
<td>X (β) 0.244, M (β) 0.169</td>
</tr>
<tr>
<td>( \beta_2 = 0.5, \beta_3 = 0.5, \beta_8 = 0.5 )</td>
<td>X (β) 0.321, M (β) 0.324</td>
<td>X (β) 0.577, M (β) 0.515</td>
</tr>
<tr>
<td>( \beta_2 = 0, \beta_3 = 0, \beta_8 = 0.5 )</td>
<td>X (β) 0.788, M (β) 0.863</td>
<td>X (β) 0.941, M (β) 0.487</td>
</tr>
</tbody>
</table>

Note: Bold coefficients are at least 1.96 their standard errors. Average bias \( \varepsilon_p \) of predicted \( \beta \) parameters:

- \( 0.1 \leq \varepsilon_p \leq 0.3 \)
- \( 0.3 \leq \varepsilon_p < 0.5 \)
- \( \varepsilon_p > 0.5 \)
Table 5  Generalized Method of Moments (GMM) and Fixed-Effects Cross-Lagged Panel-Model Regressions (FE-CLPM) of simulated data

<table>
<thead>
<tr>
<th></th>
<th>GMMS</th>
<th>FE-CLPMs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>contemporary lagged consecutive</td>
<td>consecutive + L₁</td>
</tr>
<tr>
<td></td>
<td>β̂ (se)</td>
<td>β̂ (se)</td>
</tr>
<tr>
<td>( β_2 = 0, β_3 = 0, β_8 = 0 )</td>
<td>0.375 (0.013)</td>
<td>0.221 (0.036)</td>
</tr>
<tr>
<td>( β_2 = 0, β_3 = 0, β_8 = 0 )</td>
<td>0.452 (0.019)</td>
<td>0.294 (0.038)</td>
</tr>
<tr>
<td>( β_2 = 0, β_3 = 0, β_8 = 0 )</td>
<td>0.349 (0.021)</td>
<td>0.430 (0.039)</td>
</tr>
<tr>
<td>( β_2 = 0, β_3 = 0, β_8 = 0 )</td>
<td>0.552 (0.024)</td>
<td>0.502 (0.043)</td>
</tr>
<tr>
<td>( β_2 = 0, β_3 = 0, β_8 = 0 )</td>
<td>0.375 (0.013)</td>
<td>0.237 (0.039)</td>
</tr>
<tr>
<td>( β_2 = 0, β_3 = 0, β_8 = 0 )</td>
<td>0.637 (0.014)</td>
<td>0.637 (0.014)</td>
</tr>
</tbody>
</table>

Note: Bold coefficients are at least 1.96 their standard errors. Average bias ε̄ of predicted β parameters:

- \( 0.1 < ε̄ < 0.3 \)
- \( 0.3 < ε̄ < 0.5 \)
- \( ε̄ > 0.5 \)
Table 6  Causal mediation (CM) and causal mediation regression analysis with fixed effects (CM-FE) of simulated data

<table>
<thead>
<tr>
<th></th>
<th>Causal mediation (without fixed effects)</th>
<th>Causal mediation (with fixed effects)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>contemporary</td>
<td>lagged</td>
</tr>
<tr>
<td></td>
<td>continuous + l₁,ᵢ</td>
<td>continuous + l₁,ᵢ</td>
</tr>
<tr>
<td></td>
<td>continuous + l₁,ᵢ</td>
<td>continuous + l₁,ᵢ</td>
</tr>
<tr>
<td>β₁ = β₂ = β₃ = β₉ = 0</td>
<td>β̂ (sc) 0.221 0.175 0.204 0.231 0.257</td>
<td>β̂ (sc) 0.221 0.175 0.204 0.231 0.257</td>
</tr>
<tr>
<td></td>
<td>total</td>
<td>total</td>
</tr>
<tr>
<td></td>
<td>ACME 0.123 0.141 0.159 0.176 0.193</td>
<td>ACME 0.123 0.141 0.159 0.176 0.193</td>
</tr>
<tr>
<td>β₂ = 0.5,β₃ = 0,β₉ = 0</td>
<td>β̂ (sc) 0.390 0.254 0.288 0.322 0.356</td>
<td>β̂ (sc) 0.390 0.254 0.288 0.322 0.356</td>
</tr>
<tr>
<td></td>
<td>total</td>
<td>total</td>
</tr>
<tr>
<td></td>
<td>ACME 0.180 0.207 0.224 0.241 0.258</td>
<td>ACME 0.180 0.207 0.224 0.241 0.258</td>
</tr>
<tr>
<td>β₂ = 0.5,β₃ = 0.5,β₉ = 0</td>
<td>β̂ (sc) 0.430 0.294 0.328 0.362 0.396</td>
<td>β̂ (sc) 0.430 0.294 0.328 0.362 0.396</td>
</tr>
<tr>
<td></td>
<td>total</td>
<td>total</td>
</tr>
<tr>
<td></td>
<td>ACME 0.221 0.248 0.265 0.282 0.300</td>
<td>ACME 0.221 0.248 0.265 0.282 0.300</td>
</tr>
<tr>
<td>β₂ = 0,β₃ = 0.5,β₉ = 0.5</td>
<td>β̂ (sc) 0.470 0.334 0.368 0.402 0.436</td>
<td>β̂ (sc) 0.470 0.334 0.368 0.402 0.436</td>
</tr>
<tr>
<td></td>
<td>total</td>
<td>total</td>
</tr>
<tr>
<td></td>
<td>ACME 0.271 0.298 0.315 0.332 0.350</td>
<td>ACME 0.271 0.298 0.315 0.332 0.350</td>
</tr>
<tr>
<td>β₂ = 0,β₃ = 0,β₉ = 0.5</td>
<td>β̂ (sc) 0.510 0.374 0.408 0.442 0.476</td>
<td>β̂ (sc) 0.510 0.374 0.408 0.442 0.476</td>
</tr>
<tr>
<td></td>
<td>total</td>
<td>total</td>
</tr>
<tr>
<td></td>
<td>ACME 0.321 0.348 0.365 0.382 0.400</td>
<td>ACME 0.321 0.348 0.365 0.382 0.400</td>
</tr>
<tr>
<td>β₂ = 0.5,β₃ = 0.5,β₉ = 0.5</td>
<td>β̂ (sc) 0.550 0.414 0.448 0.482 0.516</td>
<td>β̂ (sc) 0.550 0.414 0.448 0.482 0.516</td>
</tr>
<tr>
<td></td>
<td>total</td>
<td>total</td>
</tr>
<tr>
<td></td>
<td>ACME 0.371 0.398 0.415 0.432 0.450</td>
<td>ACME 0.371 0.398 0.415 0.432 0.450</td>
</tr>
</tbody>
</table>

Note: Bold coefficients are at least 1.96 their standard errors. Average bias of predicted β parameters:

- 0.1 ≤ β̂ ≤ 0.3
- 0.3 ≤ β̂ ≤ 0.5
- β̂ > 0.5
- β̂ < -0.1
- 0.3 ≤ β̂ ≤ -0.5
- β̂ < -0.5
In contrast to the OLS approach, the estimates of the FE approach (see right panel of Table 4) tend to be downwardly biased (though its bias is generally smaller compared to OLS). While some of the estimates are significantly negative although the respective \( \beta \) coefficients have been set to zero, several predicted values of both \( \beta_2 \) and \( \beta_3 \) get remarkably close to the generated ones in the consecutive scenario (C) without modeling an effect of \( L.Y \) (which had been set in the DGP, though) in particular. Moreover, on the one hand, the FE approach does not stand at risk to erroneously predict a mediation effect that has not been introduced in the DGP (Appendix A, Figure A2a). On the other hand, however, once a mediation effect is considered in the DGP, it is correctly identified by scenario C) only (Appendix A, Figure A2b).

The average bias of the GMM approach is even smaller compared to the FE approach (see left panel of Table 5). Note that the contemporary scenario (A) of the GMM approach is a replication of the corresponding OLS approach modeled as a special case of GMM – which is why the respective point estimates are almost identical to the contemporary scenario from the OLS approach (left panel in Table 4); with smaller standard errors, though. While there is some amount of downward bias in the effect of \( X \) in the lagged scenario (B), the consecutive scenario (C) in particular performs very well to detect the coefficients modeled in the DGP (although their corresponding confidence intervals still overlap in case of \( \beta_2 = \beta_3 = \beta_8 = .5 \); see Appendix A, Figure 3b). Interestingly, the consecutive scenario which controls for the lag of \( Y \) (D) is also slightly biased downwardly once \( \beta_3 \) has been set to .5.

The FE-CLPM approach (right panel of Table 5) yields results that are, on average, similarly accurate as the ones produced by the GMM approach – but with a few differences that deserve to be carved out: First, while most parameter estimates from the contemporary scenario (A) of the GMM approach (which is equivalent to the one by the OLS approach) are upwardly biased, most parameter estimates from the contemporary scenario of the FE-CLPM approach are downwardly biased. Second, the downward bias in the lagged scenario (B) of the FE-CLPM approach is comparable to the one of the GMM approach. Third, in the consecutive scenarios (C) and (D), the FE-CLMPs correctly identify both the effects of \( X \) and \( M \) on \( Y \) as well as the mediation effects once they have been modeled in the DGP (also see Appendix A, Figure A4). Fourth, as an advantage to the GMM approach, the predicted coefficients from the FE-CLMP approach are less sensitive towards the specification of the first lag of \( Y \) in the estimation process.

In the CM models without fixed effects (left panel of Table 6), the parameter estimates can be biased in both directions: On the one hand, in case of \( \beta_2 = \beta_3 = \beta_5 = 0 \), significant positive effects are predicted for all parameters (including the ACME) although they have been absent in the DGP. On the other hand, in case of \( \beta_2 = .5 \) and \( \beta_8 = .5 \), the direct effect of \( X \) on \( Y \) is notably underestimated within
all scenarios, while the effect of $M$ is still overestimated.\footnote{For ease of interpretation, recall that the total effect of $X$ on $Y$ ($TE_{XY}$) is computed as follows: $TE_{XY} = \beta_2 + (\beta_3 \cdot \beta_8)$.} The coefficient plots of the CM models without fixed effects are displayed in Figure A5 of Appendix A.

Finally, in the CMFE models (right panel of Table 6), most predicted parameters suffer from a considerable downward bias. For instance, in case of $\beta_2 = \beta_3 = \beta_8 = 0$, all parameter estimates of the \textit{contemporary} scenario (A) are negative. While the other scenarios correctly identify the above effects to be absent, for other values of $\beta_2$, $\beta_3$ and $\beta_8$, they likewise fail to identify effects that should be present according to the DGP (i.e., the corresponding coefficients are not significant). This bias of the CMFE approach is most pronounced in case of $\beta_2 = \beta_3 = 8 = .5$. The coefficient plots of the CMFE models are displayed in Figure A6 of Appendix A. A concise summary and corresponding interpretation of all findings will be given in the conclusion section.

**Conclusion**

The aim of this paper was to provide both a theoretical foundation and an empirical examination of different statistical approaches to mediation analysis. Regarding theory, a brief sketch of the fundamentals of mechanism-based explanations set the argument of adhering to a consecutive order of predictor, mediator and outcome in mediation analysis. Having summarized the statistical fundamentals of different approaches to mediation analysis, I provided a simulation analysis of the data-generating process (DGP) which could be actively manipulated to examine differences in relative performance under different scenarios: A) all-simultaneous, B) first lag of all coefficients; C) consecutive order; D) consecutive order plus first lag of $Y$ as a predictor. Each scenario was analyzed by the following methods: OLS regressions, fixed effects (FE) regressions, generalized method of moments (GMM) regressions, causal mediation analysis without (CM) and with fixed effects (CMFE), and fixed-effects cross-lagged panel models (FE-CLPMs).

The results of the simulation study suggest that the estimates of the OLS approach are generally upwardly biased, the ones of the FE and CMFE regressions are by trend downwardly biased, and the ones of the CM models (without FEs) can be biased in both directions. In contrast, the coefficients and confidence intervals estimated by both GMM regressions and FE-CLPMs are most accurate, in particular if the structure of lags in the empirical models met the consecutive order which had been set up in the underlying DGP. Most interestingly, while the GMM approach tended to be sensitive against whether or not the first lag of $Y$ ($L.Y$) was modeled as an additional predictor (the autocorrelation of $Y$ was set
to .5 in all models), the FE-CLPMs appeared to be insensitive in this respect. As a first practical implication, FE-CLPMs could be more applicable in cases of mediation analysis where the researcher is not sure whether or not $L.Y$ should be included as a predictor. A second practical implication is that even GMM regressions and FE-CLPMs can only detect the true parameter values when the order of the DGP is met. Consequently, it is of utmost importance that researchers most carefully translate their theoretical assumptions into an empirical model with the appropriate causal order: if a researcher is theoretically convinced that the causal order of the hypothesized effect is $X_{(t-2)} \rightarrow M_{(t-1)} \rightarrow Y_t$, then naively predicting $Y_t$ by $X_t$ and $M_t$ or even by $X_{(t-1)}$ and $M_{(t-1)}$ in any applied data might yield biased results irrespective of the statistical method used.

Concerning directions for future research, one direct advancement would be to shed more light on how different values for the autocorrelation of $Y$ affect the extent to which the results of the GMM approach depend on the inclusion of $L.Y$ as an additional predictor of $Y$. A second, more challenging direction could be to consider more complex data structures (such as time nested in individuals nested in additional contexts) or modeling purposes (such as moderated mediation). As a third, related, direction, future simulation studies could manipulate different forms of observed heterogeneity (between $X$ and $Z$, $M$ and $Z$, and/or $X$ and $M$) to explore the performance of each approach to mediation under different scenarios of moderation.

All in all, analyzing various DGP scenarios by different statistical approaches to mediation analysis will yield important implications for applied researchers who aim to translate particular mechanism-based explanations in statistical mediation models.

**References**


