Sociological Methods & Research

The Future Strikes Back: Using Future Treatments to Detect and Reduce Hidden Bias

Felix Elwert, Fabian T. Pfeffer

First Published October 3, 2019 | Research Article
https://doi.org/10.1177/0049124119875958

Abstract

Conventional advice discourages controlling for postoutcome variables in regression analysis. By contrast, we show that controlling for commonly available postoutcome (i.e., future) values of the treatment variable can help detect, reduce, and even remove omitted variable bias (unobserved confounding). The premise is that the same unobserved confounder that affects treatment also affects the future value of the treatment. Future treatments thus proxy for the unmeasured confounder, and researchers can exploit these proxy measures productively. We establish several new results: Regarding a commonly assumed data-generating process involving future treatments, we (1) introduce a simple new approach and show that it strictly reduces bias, (2) elaborate on existing approaches and show that they can increase bias, (3) assess the relative merits of alternative approaches, and (4) analyze true state dependence and selection as key challenges. (5) Importantly, we also introduce a new nonparametric test that uses future treatments to detect hidden bias even when future-treatment estimation fails to reduce bias. We illustrate these results empirically with an analysis of the effect of parental income on children's educational attainment.

Keywords

causal inference, confounding, bias, directed acyclic graphs, future treatments

Introduction

Hidden bias from unobserved confounding is a central problem in the social sciences. If unobserved variables affect both the treatment and the outcome, then conventional regression and matching estimators cannot recover causal effects (e.g., Morgan and Winship 2015; Rosenbaum 2002). One set of strategies for mitigating confounding bias that has been used in scattered contributions in sociology and economics involves *future treatments*, that is, values of the treatment that are realized after the outcome has occurred. The basic intuition behind these strategies is that the same unobserved confounder that affects the treatment variable before the outcome often also affects a future value of the treatment variable, measured after the outcome. If so, future values of the treatment are proxy measures of the unmeasured confounder and may help remove bias.

A few authors have previously appealed to this intuition and proposed a variety of different estimators. For instance, prior research has exploited future treatments in structural equation models (SEM) (Mayer 1997), used future treatments to measure and subtract unobserved bias (Gottschalk 1996), and employed them as instrumental variables (Duncan, Connell, and Klebanov 1997).[1]

We posit that future-treatment strategies hold significant promise for social science research for several reasons. First, future treatments can help detect, reduce, and even remove bias from unobserved confounding. Second, future values of the treatment are routinely available in panel data. Third, since future-treatment strategies require only that the *treatment variable* varies over time (i.e., not the outcome), they are available even when individual-level fixed-effects panel estimators are not. Fourth, since different future-treatment strategies impose different assumptions about the data-generating process (DGP), they are applicable across a wide range of different substantive settings.

In this article, we analyze several prior future-treatment strategies and propose a new and simpler, but more robust, new strategy. We discuss the conditions under which future values of the treatment can reduce or fully remove confounding bias. We also highlight the conditions under which future-treatment strategies introduce more bias than they remove. Specifically, we show that future-treatment strategies are vulnerable in two scenarios: where the outcome affects future treatment (selection) and where past treatment affects future values of the treatment (true state dependence). Yet, even when future-treatment strategies fail to reduce bias, they can still be used for detecting the presence of bias. Thus, we develop a new nonparametric test for hidden bias.

We investigate the performance of future-treatment strategies across a range of data-generating processes, and we assess their relative performance compared to regula

regression estimates without corrections for unobserved confounding. We present our analysis in two complementary formats. First, we present our analysis graphically to assist empirical researchers in determining quickly whether a future-treatment strategy is appropriate for their substantive application. Second, we assume linearity to link our graphical results to familiar regression models and to quantify biases. (Online Appendices discuss related approaches, instrumental variables estimation, and provide proofs.) Finally, we illustrate the application of future-treatment strategies with an empirical example that estimates the effect of parental income on children's educational attainment. The analysis rules out that conventional treatment effect estimates are unbiased and underlines the attractiveness of our control estimate, which implies a smaller causal effect of parental income and children's educational attainment than that yielded by traditional regression.

Preliminaries: Directed Acyclic Graphs (DAGs), Linear Models, and Identification

In this section, we describe the tools of our formal identification analyses, following Pearl (2013). Since the causal interpretation of statistical analyses is always contingent on a theoretical model of data generation, we first review DAGs to notate the assumed DGP. Second, we state Wright's (1921) rules, which link the causal parameters of the DGP to observable statistical associations (covariances and regression coefficients) in linear models.[2] Readers familiar with DAGs and Wright's rules may skip this section.

We use DAGs to notate the causal structure of the analyst's presumed DGPs (Elwert 2013; Pearl 2009). DAGs use arrows to represent the direct causal effects between variables. We mostly focus on DAGs comprising four variables: a treatment, *T*, an outcome, *Y*, a future (postoutcome) value of the treatment, *F*, and an unobserved variable, *U*. In keeping with convention, we assume that the DAG shows all common causes shared between variables, regardless of whether these common causes are observed or unobserved. For example, in the DAG $T \leftarrow U \rightarrow Y$, *U* represents the unobserved common cause between *T* and *Y*.

DAGs empower the analyst to determine whether the observed association (e.g., a regression coefficient) between treatment and outcome identifies the causal effect or is biased. The observed association between treatment and outcome is said to *identify* the causal effect if the only open path connecting treatment and outcome is the causal pathway, $T \rightarrow Y$. The association between treatment and outcome is spurious, or biased for the causal effect, if at least one open path does not trace the causal pathway (e.g.,

$T \leftarrow U \rightarrow Y$). Whether a path is open (transmits association) or closed (does not transmit association) depends on what variables are, or are not, controlled in the analysis, and whether the path contains a collider variable. A central rule of working with graphical models states that a path is closed if it contains an uncontrolled collider or a controlled noncollider and is open otherwise. A collider is a variable that receives two inbound arrows, such as *C* in $A \rightarrow C \leftarrow B$ (Elwert and Winship 2014).

For most of this article, we assume a linear DGP with homogenous (constant) effects, the conventional workhorse of social science. The assumption of linearity may not always be terribly realistic, but it has the advantage of convenience, as it links DAGs directly to ordinary least squares (OLS) regression and conventional SEM methodology. Under linearity and homogeneity, DAGs become linear path models, and every arrow in a linear path model is fully described by its *path parameter*, *p*, which quantifies its direct causal effect. Since path parameters are causal effects, they cannot be observed directly.[3] Later, we also consider fully nonparametric models when developing a new test for unobserved confounding.

We work with standardized variables (zero mean and unit variance) throughout for ease of exposition. Standardized path parameters cannot exceed 1 in magnitude.[4] To prevent model degeneracies, we assume that all path parameters lie strictly inside the interval $-1 < p < 1$ and differ from zero, $p \neq 0$. We discuss implications of standardization when they make a practical difference below. Wright's (1921) path rules link the unobserved path parameters of the presumed linear DGP to observable covariances.

> *Wright's (1921) path rule:* The marginal (i.e., unadjusted) covariance between two standardized variables *A* and *B*, σ_{AB} equals the sum of the products of the path parameters along all open paths between *A* and *B*.

That is, to calculate the marginal covariance between two variables *A* and *B*, first compute the product of the path parameters for each of the open paths between *A* and *B* and then sum these products across all open paths. Next, we link the marginal covariances to OLS regression coefficients (with or without control variables). The OLS regression coefficient on *T* in the unadjusted regression $Y = b_{YT}T + u$ with standardized variables equals the marginal covariance between *Y* and *T*,

$$b_{YT} = \sigma_{YT}. \qquad 1$$

We call b_{YT} the *unadjusted OLS coefficient* on *T*. The partial regression coefficient on *T* after controlling for *F* in the regression $Y = b_{YT.F}T + b_{YF.T}F + u$ is given by,

$$b_{YT.F} = \frac{\sigma_{YT} - \sigma_{YF}\sigma_{FT}}{(1-\sigma_{FT}^2)}. \quad 2$$

We call $b_{YT.F}$ the *F*-adjusted coefficients on *T*. Analogously, the *T*-adjusted coefficient on *F* is given by,

$$b_{YF.T} = \frac{\sigma_{YF} - \sigma_{YT}\sigma_{FT}}{(1-\sigma_{FT}^2)}. \quad 3$$

We omit observed control variables (other than *F*) from the analysis because they do not contribute to intuition. All of our results generalize to the inclusion of pretreatment control variables.[5]

Putting these elements together, the subsequent analyses proceed in four steps. First, we draw the DAG for the DGP we wish to analyze. Second, we use Wright's rule to express the marginal covariances between observed variables in terms of the true path parameters. Third, we plug these covariances into equations (1)–(3) to obtain the regression coefficients. Finally, we investigate whether any of these regression coefficients, or functions of regression coefficients, equal (or "identify") the desired causal effect of the treatment on the outcome, and we quantify possible biases.

The Problem: Unobserved Confounding

Figure 1 highlights the problem of unobserved confounding and illustrates our running example. The DAG shows the DGP for an observational study to estimate the total causal effect of a treatment, *T* (e.g., parental income), on an outcome, *Y* (e.g., children's years of completed education). Since treatment is not randomized, the effect of *T* on *Y* is likely confounded by one or more factors, *U*, that jointly affect *T* and *Y* (e.g., parental ambition). If so, the unadjusted association between *T* and *Y* will be biased for the causal effect of *T* on *Y*, because the association will be a combination of the association transmitted along the open causal path $T \rightarrow Y$ and the open noncausal path $T \leftarrow U \rightarrow Y$. (If all confounding variables *U* are measured, then controlling for them removes all bias by closing the noncausal path $T \leftarrow U \rightarrow Y$.) Henceforth, we assume that at least one confounding factor, *U*, is unobserved. This mimics the main predicament of most observational studies in the social sciences. If Figure 1 represents a linear and homogenous DGP, then, by equation (1) and Wright's path rule, the unadjusted OLS regression coefficient on *T* equals

$$b_{YT} = \sigma_{YT} = b + ac. \quad 4$$

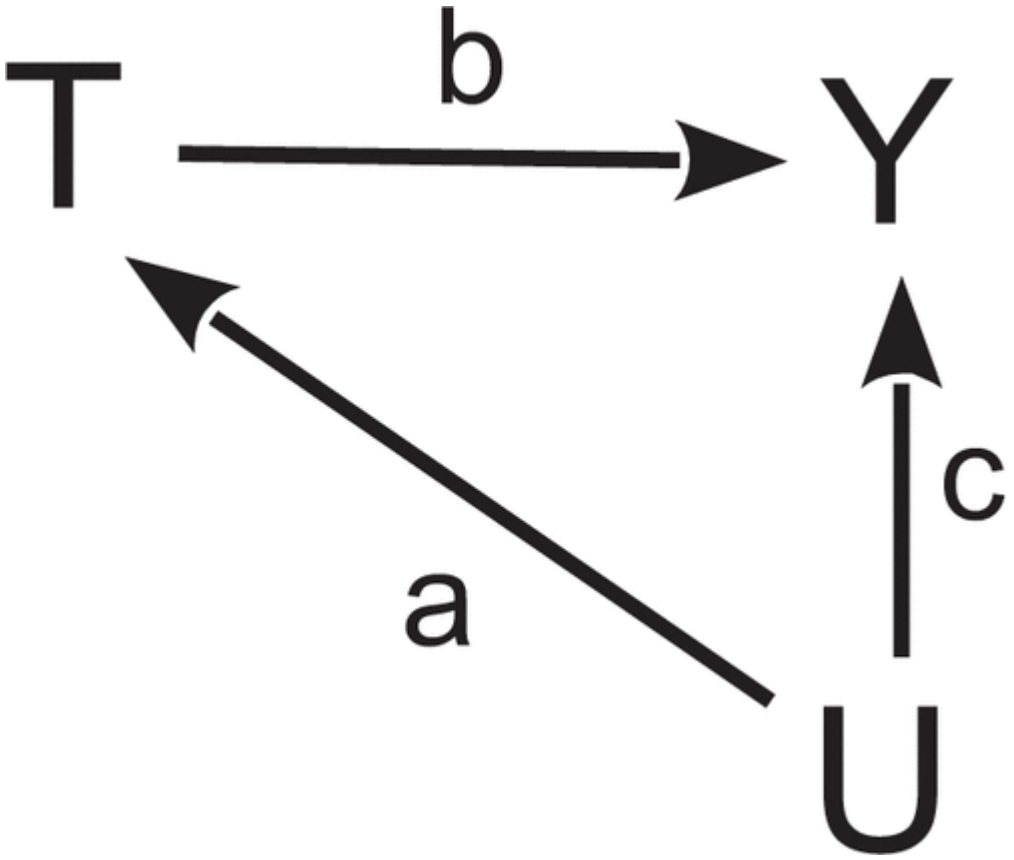

Figure 1. Directed acyclic graph for an observational study of parental income, *T*, on children's years of education, *Y*, with unobserved confounder(s), *U*, for example, parental ambition.

This regression coefficient is obviously biased for the true causal effect of *T* on *Y*, *b*. The bias equals $B_{\mathrm{OLS}} = b_{YT} - b = ac$ and increases in magnitude with the effects $U \rightarrow T$, *a* and $U \rightarrow Y$, *c*. Removing this bias from unobserved confounding is the central task of observational causal inference in the social sciences.

Strategies of Bias Correction With Future Treatments

Future treatments can be used to reduce and even fully remove bias from unobserved confounding, depending on both the analytic strategy (e.g., the chosen regression specification) and the assumptions of the DGP. In this section, we introduce two future-treatment strategies under the assumptions of the DGP shown in Figure 2. This model represents a best-case scenario for future-treatment strategies and is commonly assumed in the literature (e.g., Mayer 1997). The model assumes that the causal effect of *T* on *Y* is confounded by one or more unobserved variables, *U*, and that all unobserved factors, *U*, that confound *T* and *Y* also affect the future value of the treatment, *F*. In other words, *F* is a proxy measure for the unobserved confounder *U*.[6]

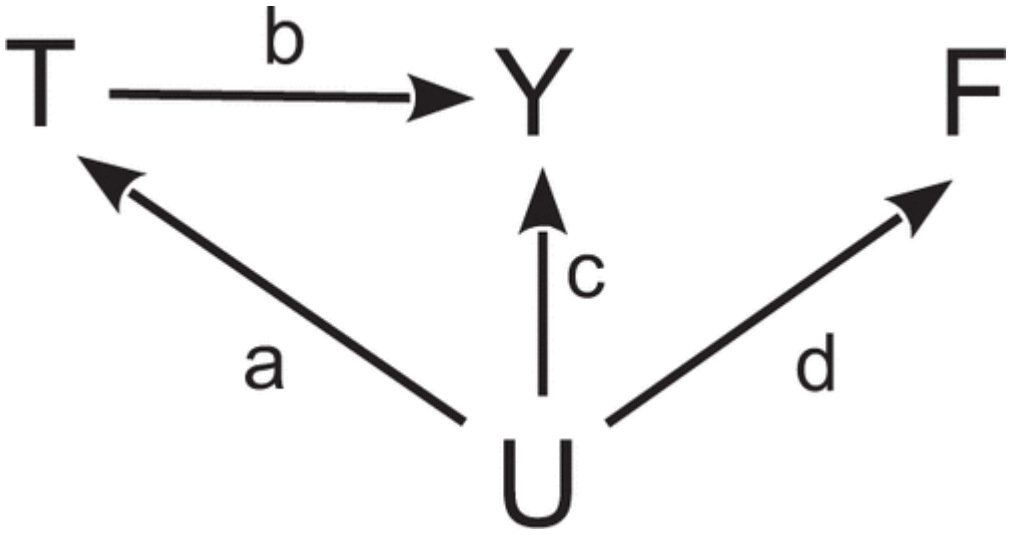

Figure 2. A confounded study where the future value of the treatment, *F*, is a proxy for the unobserved confounder(s), *U*.

The assumption that all unobserved confounders of *T* and *Y* also affect *F* is central for future-treatment strategies. Because the assumption cannot be tested empirically, it has to be defended on theoretical grounds. In many applications, it is eminently credible. For example, if parents' unmeasured ambition, *U*, affects parental income, *T*, prior to the child completing education, *Y*, it likely also affects parental income after the child has completed education, *F*.

Control Strategy of Future Treatments

Most future-treatment strategies in one way or another exploit the fact that *F* is a proxy for *U*. Here, we propose a simple estimator that exploits this fact directly: Since *F* is a proxy that carries information about *U*, controlling for *F* in the regression $Y = b_{YT.F}\,T + b_{YF.T}\,F + u$ partially controls for *U* and hence reduces bias in the treatment-effect estimate. We call the strategy of bias reduction by outright controlling for *F* the *control strategy* of future treatments.

> *Definition 1 (control-strategy estimator):* The control-strategy estimator, b_C, for the causal effect of *T* on *Y*, *b*, is given by the *F*-adjusted regression coefficient on *T*,
>
> $$b_C = b_{YT.F} = \frac{\sigma_{YT} - \sigma_{YF}\sigma_{FT}}{(1-\sigma_{FT}^2)}. \qquad 5$$

Result 1 evaluates the control-strategy estimator for data generated by the DGP in Figure 2:

> *Result 1 (bias of the control-strategy estimator in the best case):* In data generated by the DGP in Figure 2, the control-strategy estimator evaluates to:

$$b_C = b + ac\frac{(1-d^2)}{(1-a^2d^2)} = b + B_{\text{OLS}}M_C. \qquad 6$$

Clearly, the control estimator remains biased because $b_C \neq b$. Result 2, however, states that the control-strategy estimator always improves on the unadjusted OLS estimator.

> *Result 2 (strict bias reduction of the control-strategy estimator in the best case)*: In data generated by the DGP in Figure 2, the control-strategy estimate is strictly less biased than the unadjusted OLS estimate.

To see this, note that the control-strategy estimator multiplies the unadjusted OLS bias, $B_{\text{OLS}} = ac$, by the factor $M_C = \frac{(1-d^2)}{(1-a^2d^2)}$, which we call the bias multiplier of the control strategy. Since all path parameters are standardized, the control-bias multiplier is always $0 < M_C < 1$ and hence deflates the unadjusted OLS bias, $|B_{\text{OLS}}M_C| < |B_{\text{OLS}}|$. Strict bias reduction is the key advantage of the control strategy.

Figure 3 illustrates bias reduction in the control-strategy estimator compared to the unadjusted OLS estimator by graphing the absolute value of the bias multiplier of the control strategy, $|M_C|$ (dashed line), against the horizontal reference line of no bias reduction, $|M| = 1$, as a function of the strength of the effect of *U* on *F*, *d*, for a moderately strong effect of *U* on *T*, $a = .4$.[7] Clearly, the control-bias multiplier $|M_C|$ is always between 0 and 1 and hence guarantees bias reduction regardless of sign or size of the path parameters.

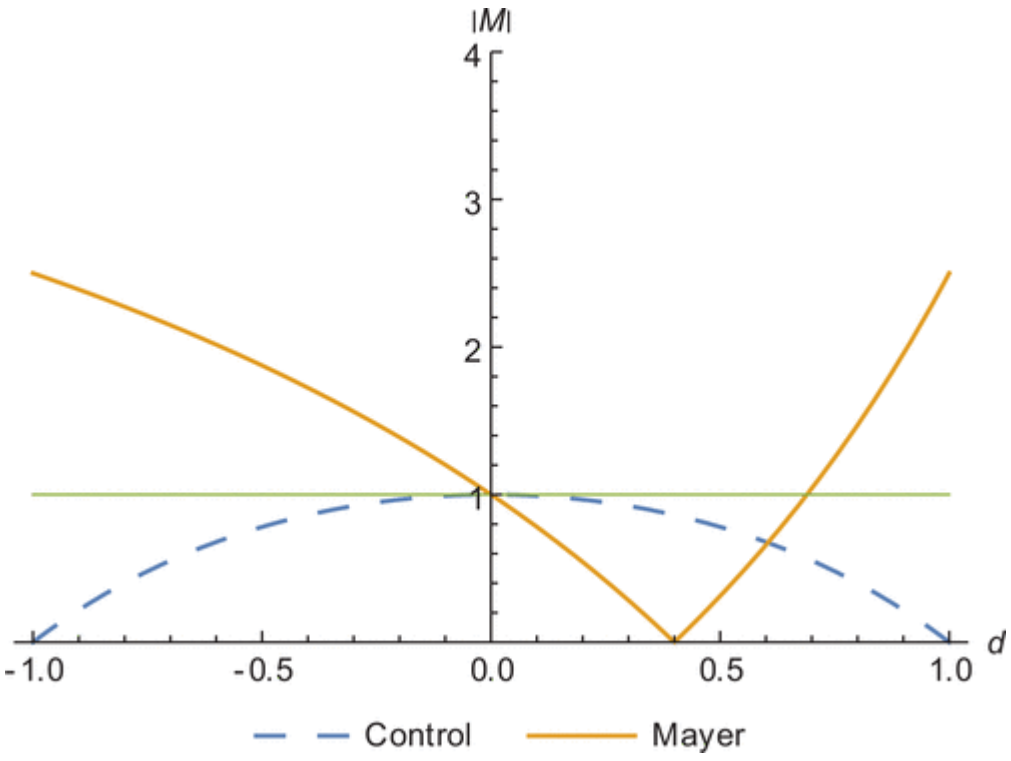

Figure 3. Absolute bias multiplier, $|M|$, for the control estimator (dashed line) and Mayer's estimator (solid line) as a function of the effect, $U \to F$, *d*. $|M| = 1$ indicates no change compared to the unadjusted OLS bias. $|M| > 1$ indicates bias amplification, $|M| < 1$ indicates bias reduction. Graphed for a moderate effect $U \to T$, $a = .4$.

The stronger the effect of *U* on *F*, $|d|$, the more bias is removed. This makes intuitive sense: the stronger the effect of *U* on *F*, the better *F* proxies for *U*. In the extreme case, where *F* is perfectly determined by U, $|d| = 1$, controlling for *F* amounts to controlling for *U* itself, thus removing all bias, such that $b_C = b$.

The control strategy gives empirical researchers a straightforward tool for reducing bias from unobserved confounding. All it takes is adding *F* as a regressor to the regression of *Y* on *T*. Controlling for *F* would also work if researchers additionally included pretreatment covariates, *X*, in the regression. The bias formulas would have to be adjusted somewhat, but the logic would remain the same. To return to our running example, under the model assumptions of Figure 2, bias in the estimated effect of parental income measured before children complete education would be strictly reduced by controlling for future parental income measured after children complete their education.

Mayer's Strategy

Mayer (1997) takes a different approach to bias reduction with future treatments. Instead of simply controlling for *F* in a regression model, she solves the structural equations of the DGP in Figure 2 under the additional assumption that the unobserved confounder, *U*, affects the future treatment, *F*, exactly like it affects the treatment, *T*, $a = d$. This assumption may be defensible in some circumstances. In our running example, one might hypothesize that parental ambition is relatively time-invariant and affects parental income, *T* and *F*, similarly at all times. Under the assumption that $a = d$, the three observable covariances between *T*, *Y*, and *F* in Figure 2, by Wright's rule, are functions of three unknown path parameters,

$$\begin{aligned} \sigma_{YT} &= b + ac \\ \sigma_{YF} &= a^2 b + ac \\ \sigma_{TF} &= a^2. \end{aligned} \tag{7}$$

This system is solved uniquely for the desired causal effect,[8]

$$\frac{\sigma_{YT} - \sigma_{YF}}{1 - \sigma_{TF}} = \frac{b + ac - ac - a^2 b}{1 - a^2} = \frac{b(1 - a^2)}{1 - a^2} = b. \tag{8}$$

Definition 2 (Mayer's [1997] estimator): Mayer's estimator for the causal effect of *T* on *Y*, *b*, is given by,

$$b_M = \frac{\sigma_{YT} - \sigma_{YF}}{1 - \sigma_{TF}}. \quad 9$$

The advantage of Mayer's estimator is that it removes all bias under the assumptions that the data are generated as in Figure 2 and that *U* affects *T* exactly as it affects *F*, $a = d$. However, when *U* affects *T* and *F* differently, $a \neq d$, then Mayer's estimator has two disadvantages. First, as Mayer (1997) notes, the estimator is biased.[9] Result 3 evaluates the bias.

Result 3 (bias of Mayer's [1997] estimator in the best case): In data generated by the DGP of Figure 2, Mayer's estimator evaluates to

$$b_M = b + ac\frac{a-d}{a-a^2d} = b + B_{\text{OLS}}M_M. \quad 10$$

Second, in contrast to the control-strategy estimator, Mayer's estimator can increase the unadjusted OLS bias, as shown in result 4.

Result 4 (bias amplification in Mayer's [1997] estimator in the best case): In data generated by the process of Figure 2, Mayer's estimator increases bias compared to the unadjusted OLS estimate when $|M_M| = |\frac{a-d}{a-a^2d}| > 1$. This occurs (1) when $\frac{a}{d} < 0$ or (2) when $|\frac{2a}{1+a^2}| < |d|$.

In other words, bias amplification occurs either (1) when *U* affects *T* and *F* in opposite directions or (2) when *U* affects *T* and *F* in the same direction, but the magnitude of the effect $U \to F$, *d*, substantially (roughly more than twice) exceeds the magnitude of the effect $U \to T$, *a*.

The solid line in Figure 3 illustrates bias reduction and bias amplification of Mayer's estimator by graphing the absolute value of the bias multiplier, $|M_M|$ across values of *d* for $a = .4$. When *a* and *d* share a sign (here, $d > 0$) and *d* is not much larger than *a*, then $|M_M| < 1$, and Mayer's estimator reduces bias. But if *a* and *d* have opposite signs (here, $d < 0$), or if $d \gg a$, then $|M_M| > 1$ and Mayer's estimator amplifies the unadjusted OLS bias.

The possibility of bias amplification in Mayer's estimator has not been noted previously. Whether bias amplification occurs depends on the empirical setting and must be carefully evaluated based on subject matter knowledge. We believe that bias amplification can be excluded in many settings. First, in many applications, *U* will not affect *T* and *F* in op

directions. In our example, it is not generally plausible that parental ambition, *U*, increases parental income early on, *T*, but decreases it later, *F*. Second, since the shared unobserved confounder *U* is by assumption a baseline characteristic that is temporally closer to *T* than to *F*, the effect of *U* on *T* will likely exceed the effect of *U* on *F*, that is, $|a| > |d|$. In our example, we are cautiously optimistic that the effect of early parental ambition is more pronounced on parent's early income, *T*, than on later income, *F*, because other determinants of income, such as experience and seniority, may grow in importance as time passes.

On the other hand, we cannot entirely rule out the possibility of bias amplification, even in our running example. Suppose, for example, that we analyze the effect of parental income on children's educational outcomes among young parents. Young parents with high ambition may still be enrolled in college and hence earn little compared to their less ambitious counterparts who already have jobs. Later, however, these highly ambitious parents may become high-earning professionals, whereas their less ambitious counterparts remain stuck in lower paying jobs. Hence, the effects $U \rightarrow T$ and $U \rightarrow F$ could have opposite signs, such that Mayer's estimator would increase rather than decrease bias. And even if the effects share the same sign, $U \rightarrow F$ may still strongly exceed $U \rightarrow T$. That is, using our example, if the returns to parental ambition compound as employees climb up the corporate ladder, early ambition may have a relatively modest effect on early income but a large effect on later income via successive promotions. If the effect of early ambition on later income sufficiently exceeds its effect on early income, then Mayer's estimator would also increase rather than decrease bias.

Implementing Mayer's Strategy as a Difference Estimator

The original presentation of Mayer's estimator required customized programming. Next, we show that Mayer's estimator straightforwardly equals the difference between two OLS regression coefficients. This enables estimation via all standard statistical software packages and provides additional intuition.

> *Definition 3 (difference estimator):* The difference estimator for the effect of *T* on *Y* is the difference between the coefficients on *T* and *F* in the regression $Y = b_{YT.F}\,T + b_{YF.T}\,F + u$,
>
> $$b_D = b_{YT.F} - b_{YF.T} = \frac{\sigma_{YT} - \sigma_{YF}\sigma_{FT}}{(1-\sigma_{FT}^2)} - \frac{\sigma_{YF} - \sigma_{YT}\sigma_{FT}}{(1-\sigma_{FT}^2)}. \qquad 11$$
>
> *Result 5 (equivalence of the difference and Mayer's estimators):*

$$b_D = \frac{\sigma_{YT} - \sigma_{YF}\sigma_{FT}}{(1-\sigma_{FT}^2)} - \frac{\sigma_{YF} - \sigma_{YT}\sigma_{FT}}{(1-\sigma_{FT}^2)}$$
$$= \frac{(1+\sigma_{FT})(\sigma_{YT} - \sigma_{YF})}{(1+\sigma_{FT})(1-\sigma_{FT})} = \frac{(\sigma_{YT} - \sigma_{YF})}{(1-\sigma_{FT})} = b_M. \square$$

The equivalence between Mayer's estimator and the difference estimator holds for all DGPs—not just the DGP of Figure 2—because the definition of the estimators only draws on empirical covariances and not on the structure of the DGP.

Equating Mayer's estimator with the difference estimator provides additional insight: The idea behind the difference estimator is to use future treatments first to measure and then to remove the spurious association between *T* and *Y*.

This fact is best appreciated by investigating the difference estimator under the assumption that the effect of *U* on *T* equals the effect of *U* on *F*, $a = d$ in data generated by Figure 2. First, the coefficient $b_{YT.F}$ is biased for *b* by the confounding path $T \leftarrow U \rightarrow Y$, less whatever part of confounding is removed by controlling for *F* (recall that *F* is a proxy for *U*). Specifically, $b_{YT.F} = b + ac\frac{(1-a^2)}{(1-a^4)}$, where $0 < \frac{(1-a^2)}{(1-a^4)} < 1$ is the deflation factor by which confounding along $T \leftarrow U \rightarrow Y$, ac is diminished by controlling for *F*. Second, the coefficient $b_{YF.T}$ captures the association flowing along the path $Y \leftarrow U \rightarrow F$, less whatever part of this association is removed by controlling for *T* (like *F*, *T* is a proxy for *U*). Specifically, $b_{YF.T} = ac\frac{(1-a^2)}{(1-a^4)}$, equals the association flowing along $Y \leftarrow U \rightarrow F$, diminished by the same deflation factor $0 < \frac{(1-a^2)}{(1-a^4)} < 1$ due to controlling for *T*. Third, clearly, $b_{YF.T}$ equals the bias in $b_{YT.F}$; hence, subtracting one from the other yields an unbiased estimate for *b*.

Expressing Mayer's estimator as a difference estimator helps explicate the properties that we claimed for it above. First, the Mayer/difference estimator removes all bias only if $a = d$, because only then does $b_{YF.T}$ exactly measure the bias in $b_{YT.F}$. More generally, by result 3, the estimator equals $b_M = b_D = b + ac\frac{a-d}{a-a^2d}$ and is biased to the extent that *a* and *d* differ. Second, if $a > d$, the estimator is biased because $b_{YF.T}$ understates the bias in $b_{YT.F}$: The association captured by the path $Y \leftarrow U \rightarrow F$ understates the bias flowing along $Y \leftarrow U \rightarrow T$. Third, if $a < d$, the estimator is biased because $b_{YF.T}$ overstates the bias in $b_{YT.F}$: The association captured by the path $Y \leftarrow U \rightarrow F$ overstates the bias flowing along $Y \leftarrow U \rightarrow T$. Fourth, if *d* is more than twice as large as *a*, then $b_{YF.T}$ may overstate the bias in $b_{YT.F}$ more than twofold, so that the Mayer/difference estimator $b_{YT.F} - b_{YF.T}$ first subtracts all bias and then more than adds it back, resulting in absolute bias amplification. Fifth, if *a* and *d* have different signs, then $b_{YF.T}$ measures the negative

of the bias in $b_{YT.F}$ such that the difference estimator $b_D = b_{YT.F} - b_{YF.T}$ adds rather than removes bias, also resulting in bias amplification.

We note that the difference estimator for future treatments has some history in social science methodology. Versions of this differencing logic are discussed by Gottschalk (1996), who explicitly uses future treatments, and by DiNardo and Pischke (1997) and Elwert and Christakis (2008), who analyze structurally similar models without future treatments. Online Appendix A evaluates Gottschalk's (1996) estimator.

The Difference Strategy of Future Treatments Is Different From Difference-in-differences (DiD)

Despite superficial similarities, the Mayer/difference strategy of future treatments differs from conventional DiD, or gain score, estimation. While both approaches assume the same qualitative causal structure for the DGP, shown in Figure 2, they impose different parametric constraints on this structure and hence derive different estimators. Mayer's approach interprets *F* as a future (postoutcome) value of the treatment and assumes that *U* affects *T* and *F* equally, $a = d$. By contrast, DiD interprets *F* as a lagged (pretreatment) value of the outcome and assumes that *U* affects *Y* and *F* equally, $c = d$ (Kim and Steiner 2019). As a result of these different constraints on the path parameters, the two methods employ different estimators. As is easily verified against the graph, with *F* as future treatment, the spurious association between *Y* and *T* is measured and removed by the conditional covariance between *Y* and *F* given *T*. Hence, the Mayer/difference estimator is $b = b_{YT.F} - b_{YF.T}$. With *F* as lagged outcome, the spurious association between *Y* and *T* equals the marginal covariance between *F* and *T*, and the DiD estimator is $b = b_{YT} - b_{TF}$.

Choosing Between Future-treatment Estimators

Next, we compare the performance of the two future-treatment strategies and provide guidance for choosing between them. We continue to assume that the data are generated by the DGP of Figure 2. Obviously, maximally cautious analysts should always prefer the control estimator, because, in contrast to the Mayer/difference estimator, it guarantees bias reduction when the data are produced by the DGP in Figure 2 regardless of the relative size of the path parameters. Bias reduction with the control estimator, however, is often quite modest. For most values of the effect $U \to T$, *a*, the control estimator will remove less than half of the unadjusted OLS bias unless the effect $U \to F$, *d* is large, $d > 0.7$. In many cases, Mayer's estimator will thus remove more bias than the control estimator.

Analysts can sometimes decide between the two future-treatment estimators by comparing the relative positions of the unadjusted OLS, control, and Mayer/difference estimates. Figure 4 illustrates the decision process. Since the control estimate, in expectation, is closer to the true treatment effect than is the unadjusted OLS estimate, the difference between the control estimate and the unadjusted OLS estimate reveals the direction of the unadjusted OLS bias. For example, if the unadjusted OLS estimate is $b_{\text{OLS}} = .5$ and the control estimate is $b_C = .3$, then the true treatment effect should be no larger than the control estimate, $b \leq .3$. The first decision rule thus states that if the control and Mayer/difference estimates change the unadjusted OLS estimate in different directions (Figure 4, scenario 1), then the analyst should choose the control estimate as bias reducing and eschew the Mayer/difference estimate as bias increasing. Second, the control estimator is preferred as long as the Mayer/difference estimator does not differ more strongly from the unadjusted OLS estimator in the same direction. For example, if the unadjusted OLS estimate is $b_{\text{OLS}} = .5$, the control estimate is $b_C = .3$, and the Mayer/difference estimate is $b_M = .4$ (Figure 4, scenario 2), then the control estimate is preferred.

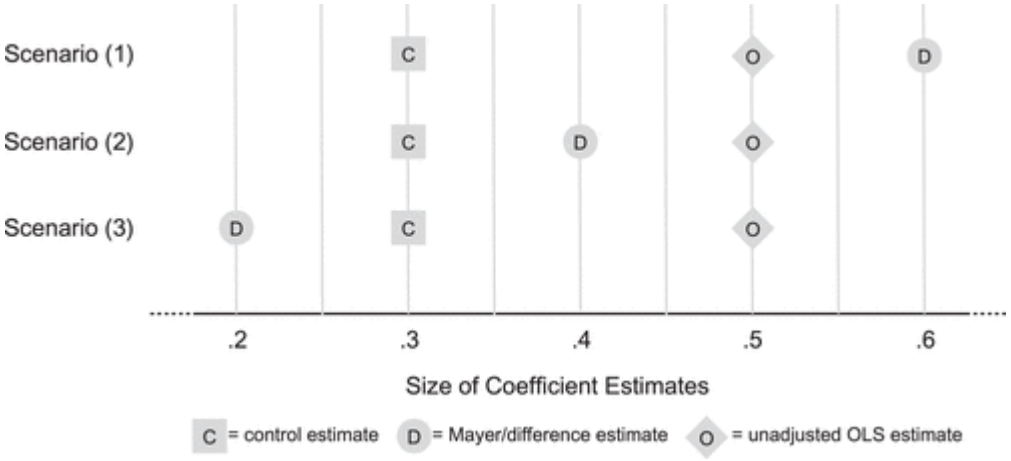

Figure 4. Illustration of the heuristic for choosing between estimates. The relative position of the control (*C*), difference (*D*), and unadjusted OLS (*O*) estimates can help the analyst decide between alternative estimates. In data generated by Figure 2, the location of the control estimate indicates the direction of unadjusted OLS bias (in this example, upward bias). In scenarios (1) and (2), the control estimate is preferred. In scenario (3), additional assumptions are needed to decide between the control and difference estimates.

If the control and Mayer/difference estimators change the unadjusted OLS estimate in the same direction, but the Mayer/difference estimator is farther away from the unadjusted OLS estimate than is the control estimate (Figure 4, scenario 3), then it does not follow that the Mayer/difference estimator is automatically preferred. For example, with $b_{\text{OLS}} = .5$, $b_C = .3$, and $b_M = .2$, the true effect could be closer to either the control estimate or the Mayer/difference estimate. Thus, the analyst would require additional knowledge about the relative size of effects $U \rightarrow T$, *a* and $U \rightarrow F$, *d* to decide betwe

the control and Mayer/difference estimates. Two rules from result 4, illustrated in Figure 3, help with this decision. First, if the analyst can argue that *a* and *d* share the same sign and that the magnitude of *a* does not considerably exceed the magnitude of *d*, then the analyst should choose the Mayer/difference estimator because it is expected to remove more bias than the control estimator. Second, if $|d| \gg |a|$ or if *d* and *a* have opposite signs, then the Mayer/difference estimator is expected to increase the unadjusted OLS bias, and the analyst should chose the control estimator.

Challenges to Bias Correction With Future Treatments

The DGP of Figure 2, analyzed so far, provides a best-case scenario for future-treatment strategies to reduce confounding bias in unadjusted OLS regressions because it guarantees bias reduction for the control estimator and even complete bias removal with the Mayer/difference estimator if $a = d$. In this section, we explain that both future-treatment strategies can increase bias in the presence of either (1) true state dependence, where past treatment causally affects future treatment, or (2) selection, where the outcome causally affects future treatment, or both. We demonstrate this failure by showing that both future-treatment strategies can produce bias even when the unadjusted OLS estimate is unconfounded and hence unbiased. With either true state dependence or selection, choosing the best future-treatment strategy becomes a matter of carefully weighing prior knowledge about the underlying path parameters in the DGP.

True State Dependence: When Treatment Affects Future Treatment

Past and future values of the treatment are typically correlated over time. One reason for this association could be mutual dependence of *T* and *F* on the unmeasured confounder *U* along the path $T \leftarrow U \rightarrow F$, as in Figure 2, which would justify the future-treatment strategies discussed above. Another reason for a correlation between *T* and *F* could be true state dependence, where past states of the treatment cause future states of the treatment (Bates and Neyman 1951; Heckman 1981a, 1981b).[10] True state dependence is captured by the arrow $T \rightarrow F$ in Figure 5. Sociologists are amply familiar with cumulative advantage and cumulative disadvantage as important special cases of true state dependence. DiPrete and Eirich (2006:272) explain that cumulative (dis)advantage "becomes part of an explanation for growing inequality when current levels of accumulation have a direct causal relationship on future levels of accumulation." For instance, individuals with higher incomes accumulate more financial assets, which in turn generate asset income returns that grow at a higher rate than earnings (Piketty 2014). In this example, as in others, a causal story of true state dependence involves a causa

mediator (here: income → financial asset acquisition → income), and the strength of state dependence may be quite limited (e.g., for most people, asset income plays no appreciable role in determining total income).

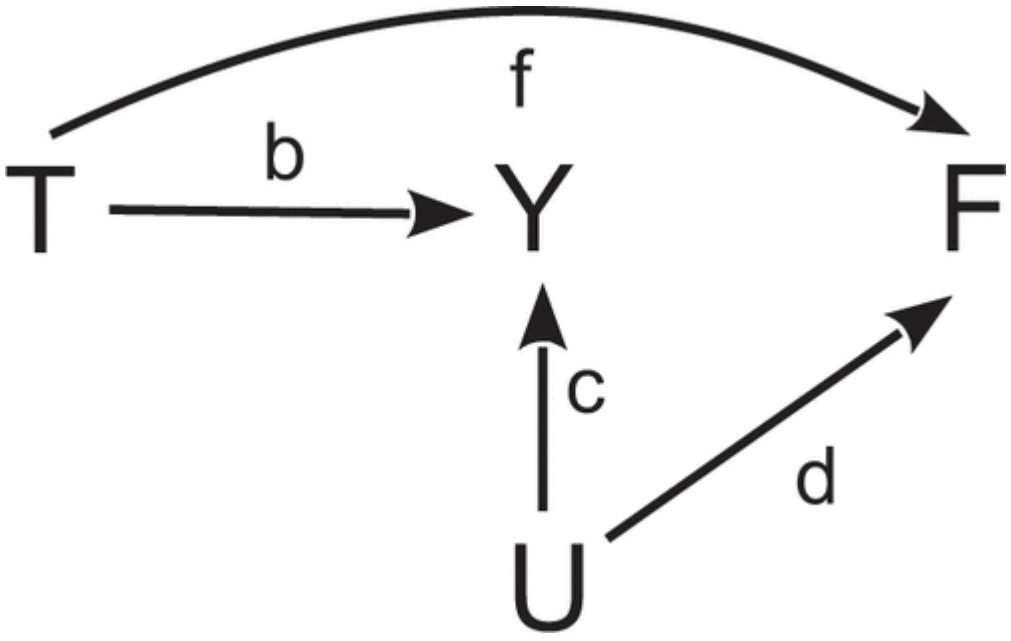

Figure 5. An unconfounded study with true state dependence of treatment, $T \rightarrow F$.

To isolate the problem of true state dependence, we first analyze the performance of future-treatment strategies when the effect of *T* on *Y* is not confounded (no arrow $U \rightarrow T$), as in Figure 5. Here, the marginal association between *T* and *Y* identifies the causal effect of *T* on *Y* because the causal effect $T \rightarrow Y$ is the only open path between them. Hence, the unadjusted OLS estimate equals the true causal effect, $b_{\text{OLS}} = b_{YT} = b$, and the unadjusted OLS estimator is unbiased.

Future-treatment strategies are vulnerable to true state dependence because needlessly controlling for *F* would introduce bias. Since *F* is a collider variable on the noncausal path $T \rightarrow F \leftarrow U \rightarrow Y$, controlling for *F* in the regression of *Y* on *T* opens this path and induces a spurious association between *T* and *Y*. Controlling for *F* would therefore create bias where none existed before. This intuition is confirmed algebraically using Wright's rules. The control estimator for data generated by Figure 5 (with true state dependence but without confounding) evaluates to

$$b_C = b_{YT.F} = b - \frac{cdf}{(1-f^2)}. \qquad 13$$

Note that the control estimator in this scenario is biased even though the unadjusted OLS estimator is not. As expected, the bias in the control estimator under true state dependence is a function of the path parameters on the noncausal path $T \rightarrow F \leftarrow U$

f, d, and c. The bias in b_C increases with the strength of confounding between *Y* and *F*, cd, in the numerator of the bias; and the bias increases especially strongly with the strength of state dependence, *f*, which increases the numerator and decreases the denominator of the bias. The Mayer/difference estimator for data generated by Figure 5, with true state dependence and without confounding, evaluates to

$$b_{M/D} = b - \frac{cd(f-1)}{(1-f^2)}. \qquad 14$$

The Mayer/difference estimator is also biased in this scenario, even though the unadjusted OLS estimator is not. Comparing equations (13) and (14) shows that true state dependence introduces less bias into the control-strategy estimator than into the Mayer/difference estimator, unless true state dependence is strongly positive, $f > 0.5$. In sum, both future-treatment estimators can increase bias under true state dependence, but the control estimator will be less biased as long as true state dependence is not too large.

Next, we analyze the empirically more interesting DGP of Figure 6, which combines Figure 2 with Figure 5 to form a scenario of true state dependence with unobserved confounding. Here, *U* is a confounder of *T* and *F*, which motivates the use of *F* as a proxy control to reduce bias in the unadjusted OLS estimator, but *T* also directly causes *F* via true state dependence, thus introducing bias into both future-treatment estimators. Without further restrictions, the analytic expressions for the control and difference estimators are unwieldy and scarcely informative (not shown). Depending on the exact parameter constellation, both future-treatment strategies could reduce bias or increase bias in the unadjusted OLS estimator. Hence, analysts must carefully consider existence, direction, and size of true state dependence in their empirical applications.

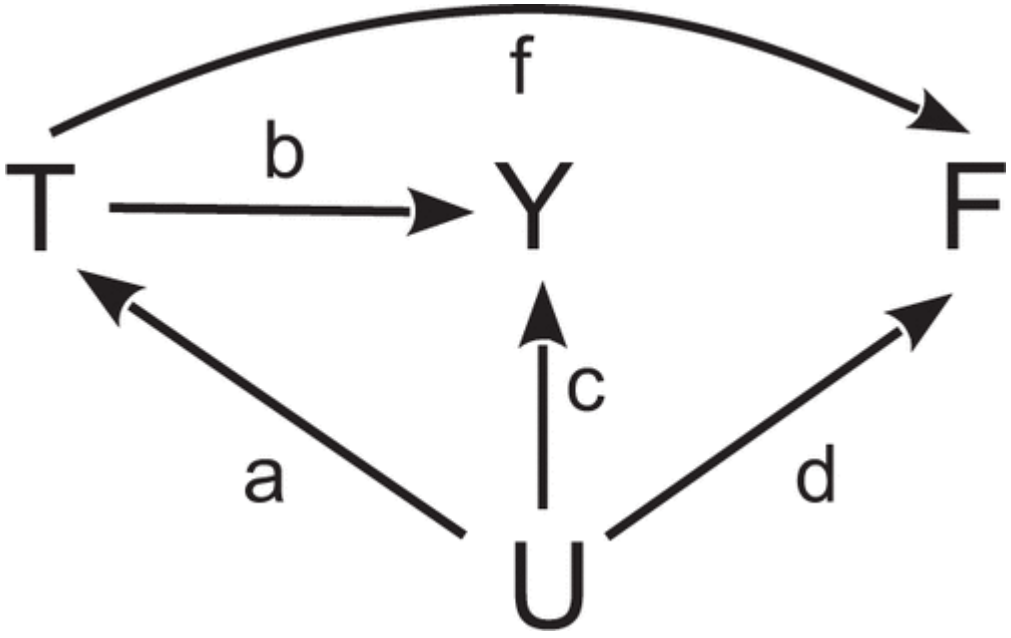

Figure 6. A confounded study with true state dependence of treatment (combination of Figures 2 and 5).

Nonetheless, future-treatment strategies remain promising if the analyst can defend certain parametric restrictions on the relative size of the path parameters. Consider, for example, the restriction that *U* affects *T* to the same extent as it affects *F*, $a = d$, as Mayer (1997) proposed for the effect of parental income on child outcomes.

Result 6 (bias of the control estimator with true state dependence): In data generated by the model in Figure 6 with the constraint $a = d$, the control estimator evaluates to

$$b_C = b + ac\frac{-f-f^2-a^2f-a^2+1}{1-(f+a^2)^2} = b + B_{\text{OLS}}R_C. \quad 15$$

Result 7 (bias of the Mayer/difference estimator with true state dependence): In data generated by the model in Figure 6 with the constraint $a = d$, the Mayer/difference estimator evaluates to

$$b_{M/D} = b + ac\frac{-f-f^2-a^2f}{1-(f+a^2)^2} = b + B_{\text{OLS}}R_M. \quad 16$$

The bias multipliers of the control and Mayer/difference estimators, R_C and R_M, are obviously closely related, though their behavior is somewhat surprising. Explorations of the parameter space (see Online Appendix E) reveal several facts, summarized in Table 1:

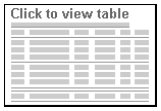

Table 1. Performance of the Control Estimator and the Mayer/Difference Estimator in the Presence of State Dependence and Assuming $a = d$.

Result 8 (relative performance of the control and Mayer/difference estimators under confounding and true state dependence): In data generated by the model in Figure 6 with the constraint $a = d$, the following five facts hold approximately:

With (often unrealistic) negative[11] state dependence, $f < 0$,

1. The Mayer/difference estimator is strictly bias reducing, $0 < R_M < 1$, and strictly dominates the performance of the control estimator, $R_M < R_C$.
2. The control estimator is strictly bias amplifying, and the bias increases as true state dependence becomes more negative.

With (often realistic) positive state dependence, $f > 0$,

(3) The Mayer/difference estimator reduces bias up to moderate positive state dependence, $0 \leq f \leq .05$ for up to moderately strong confounding, $|a| < 5$. It strictly amplifies bias above $f > 05$.

(4) The control estimator reduces bias for most values of positive state dependence, $0 \leq f \leq 0.78$, for up to moderately strong confounding, $|a| < 0.5$. It strictly amplifies bias above $f \geq 0.78$.

(5) The control estimator is less biased than the Mayer/difference estimator above $f = 0.37$, and more biased otherwise.

Table 1 underlines that true state dependence, which is a common concern in sociology, ruins the strict bias-reduction property of the control estimator. To muddy matters further, judging the size of the standardized path parameters in this scenario is not easy in practice. The difficulty is that the variances of *T* and *F* almost certainly differ when true path dependence is present. Hence, it is difficult to assert the equality of $a = d$ because equality in the standardized parameters does not imply equality of the effects of *U* on *T* and *F*, respectively, in their natural scale.[12]

Nonetheless, under realistic values of moderate positive true state dependence, both the control and the Mayer/difference estimators are likely bias reducing. And for moderate and strong positive state dependence, the control estimator outperforms the Mayer/difference estimator and remains (strongly) bias reducing as long as the effects of *U* on *T* and *F* are not too large.

Selection Bias: When the Outcome Affects Future Treatment

Selection also complicates future-treatment strategies for unobserved confounding. We say that selection is present when the outcome exerts a causal effect on the future value of the treatment, as captured by the arrow $Y \rightarrow F$ in Figure 7. Selection is a concern in many situations. For example, in a study of the effect of parental income on educational attainment, college enrollment might affect parents' income if parents adjust their labor supply to the financial needs of the child. In other scenarios, selection may be absent. For example, when studying the effect of parental income on children's test scores, it is implausible to believe that children's test scores affect future values of parental income (except, perhaps, when a child's abysmal test scores inspire a parent to quit their job to tutor the child).

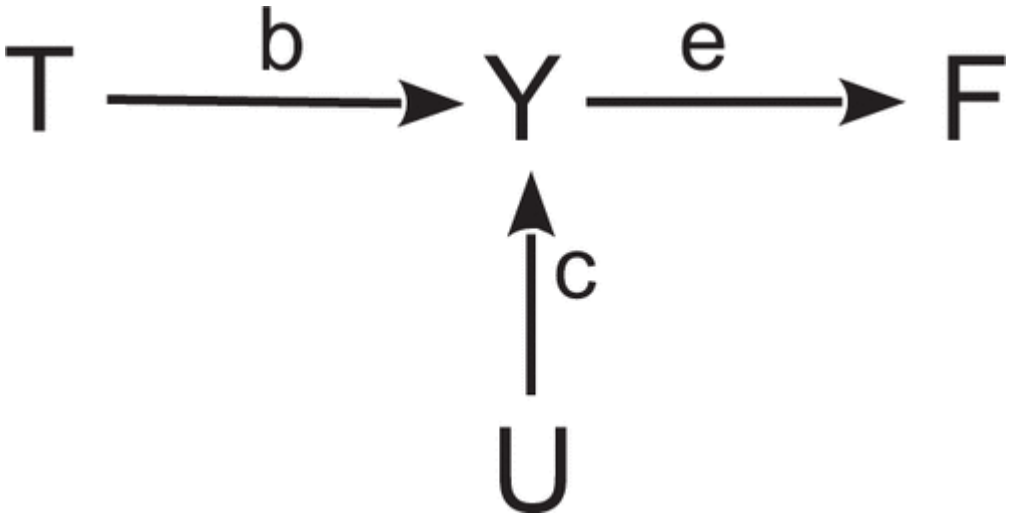

Figure 7. An unconfounded study with selection, $Y \rightarrow F$.

Figure 7 isolates the problems of selection. Since the effect of *T* on *Y* is unconfounded, the unadjusted OLS estimator again recovers the true causal effect, $b_{YT} = b$. The control and difference estimators, however, control for *F* and hence suffer selection bias because controlling for *F* amounts to selecting on the outcome.[13] The control-strategy estimator without confounding but with selection evaluates to

$$b_C = b_{YT.F} = b\frac{(1-e^2)}{1-b^2e^2} = bP_C, \qquad 17$$

and the difference strategy estimator evaluates to

$$b_{M/D} = b\frac{b-e}{b-b^2e} = bP_D. \qquad 18$$

Since $P_C \neq 1$ and $P_D \neq 1$ are pure bias terms, neither the control estimator nor the difference estimator recovers the true causal effect. It can be shown, however, that $|P_C| \ll |P_D|$; that is, selection (without confounding) introduces far less bias into the control estimator than into the Mayer/difference estimator, especially for small treatment effects $T \rightarrow Y$. Note that bias in the control and Mayer/difference estimators with selection depends on the size of the treatment effect, *b*. Finally, Figure 8 shows the empirically important scenario with both selection and confounding (combining Figures 7 and 2, respectively). The corresponding analytic expressions for the control and Mayer/difference estimators are highly nonlinear.

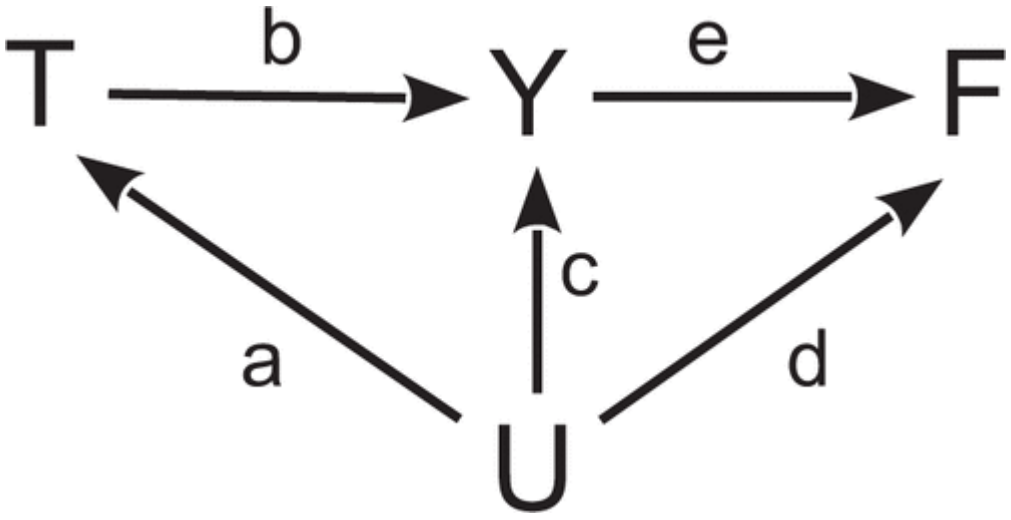

Figure 8. A confounded study with selection, $Y \rightarrow F$.

Result 9 (bias in the control estimator with selection): In data generated by Figure 8, the control estimator evaluates to

$$b_C = b_{YT.F} = \frac{(b+ac) - (bad+e+cd)\,(be+ace+ad)}{1-(be+ace+ad)^2} = b + B_{\mathrm{OLS}} S_C, \tag{19}$$

Result 10 (bias in the Mayer/difference estimator with selection): In data generated by Figure 8, the Mayer/difference estimator evaluates to

$$b_D = \frac{(b+ac) - (bad+e+cd)}{1-(be+ace+ad)} = b + B_{\mathrm{OLS}} S_D. \tag{20}$$

The bias-reduction properties of both future-treatment estimators with confounding and selection strongly depend on the underlying path parameters. Simulations (see Online Appendix) suggest that the Mayer/difference estimator is usually performing worse, and often dramatically so, than the control estimator as long as the path parameters, *p*, are not too large, $|p| < .5$. Specifically, any hint of selection, $e \neq 0$, threatens to turn the Mayer/difference estimator into a bias amplifier. By contrast, as long as selection is mild, $|e| \leq 0.3$, the control estimator remains bias reducing, though bias reduction can be small in absolute terms.[14]

Table 2 summarizes the divergent performance of the control and the Mayer/different estimator. The upshot is that for scenarios in which path parameters are at most moderately strong $(|p| \leq .5)$, the control strategy generally carries the day. Since the control strategy without selection is strictly bias reducing and only minimally biased by selection, it tends to remove some bias overall. By contrast, the Mayer/difference strategy

is strongly bias reducing without selection but can induce heavy bias with selection, and so it is to be used with caution.

Table 2. Performance of the Control Estimator and the Mayer/Difference Estimator in the Presence of Selection and Weak to Moderate Path Parameters, $|p| < .5$.

Future-treatment Tests for Unobserved Confounding

Future treatments can also be used to detect, and even formally test for, the presence of unobserved confounding between *T* and *Y*. Importantly, testing for bias is possible under substantially weaker assumptions than bias reduction or bias removal. In this section, we develop a nonparametric test for unobserved confounding via two results. (Readers uninterested in the technical details may skip directly to result 12.)

> *Definition 4:* Let *V* be an unobserved variable that directly causes treatment, *T*, $V \rightarrow T$ and is associated with a postoutcome value of the treatment, *F*, conditional on *T* and the vector of observed pretreatment covariates, ***X*** (which may be empty), $V \neg \perp F | (T, \boldsymbol{X})$.
>
> *Assumption 1*: The DGP contains at least one unobserved variable *V*.

Assumption 1 is quite minimal compared to models discussed before. First, assumption 1 requires only partial knowledge of the DGP rather than a fully articulated DAG. Second, it is nonparametric, that is, it puts no restrictions on the functional form of the effects. Third, it does not require constant effects across individuals.

> *Result 11*: Given assumption 1, independence between *F* and *Y* conditional on *T* and *X*, $F \perp Y | (T, \boldsymbol{X})$, implies the absence of unobserved confounding between *T* and *Y* conditional on **X**. (Proof in Online Appendix C)

By contraposition, result 11 says that unobserved confounding between *T* and *Y* will lead to a conditional association between *Y* and *F* (given *T* and ***X***), as long as at least one unobserved cause of *T*, *V*, is also associated with *F* (given *T* and ***X***). The scenarios of unobserved confounding thus detected include the usual situations of hidden bias, where the unobserved factors that confound *T* and *Y* are also causes of *F* (e.g., Figure 2), even with strong path dependence (e.g., Figure 6). And they also include considerably more complex situations, for example, when confounding between *T* and *Y* is only induced by control for pretreatment colliders, ***X***, and *T* shares no causes with *F* (Figure 9).

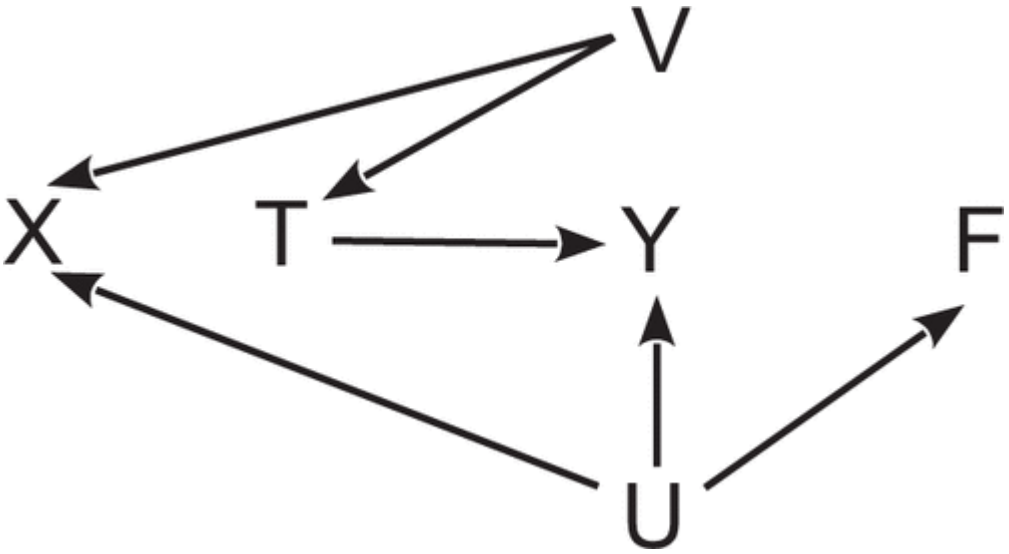

Figure 9. The causal effect of *T* on *Y* is confounded after controlling for ***X*** because conditioning on ***X*** opens the noncausal path $T \leftarrow V \rightarrow \boldsymbol{X} \leftarrow U \rightarrow Y$. Assumption 1 holds because conditioning on ***X*** also opens the path $T \leftarrow V \rightarrow \boldsymbol{X} \leftarrow U \rightarrow F$. *F* and *Y* are associated conditional on *T* and ***X*** because $Y \leftarrow U \rightarrow F$ is always open.

Result 11 thus turns the conditional independence between *F* and *Y* into an indicator for the absence of unobserved confounding between *T* and *Y*. However, result 11 does not yet justify a formal test for unobserved confounding because the result is not symmetric: Independence between *F* and *Y* (given *T* and ***X***) implies the absence of unobserved confounding, but dependence is compatible with both unobserved confounding and its absence (e.g., in Figures 10 and 11).

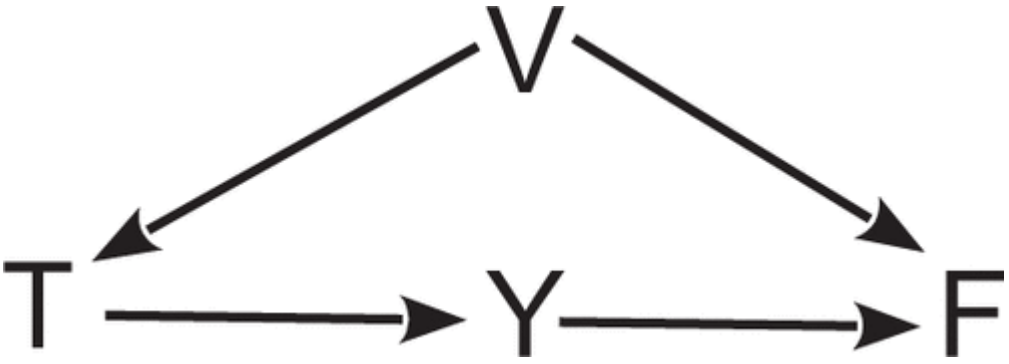

Figure 10. *F* and *Y* are associated even though there is no unobserved confounding between *T* and *Y*. Assumption 1 holds because $T \leftarrow V \rightarrow F$ is always open. *Y* and *F* are associated because $Y \rightarrow F$ is always open. *T* and *Y* are unconfounded because *F* is an unconditioned collider on the only noncausal path between them, $T \leftarrow V \rightarrow F \leftarrow Y$.

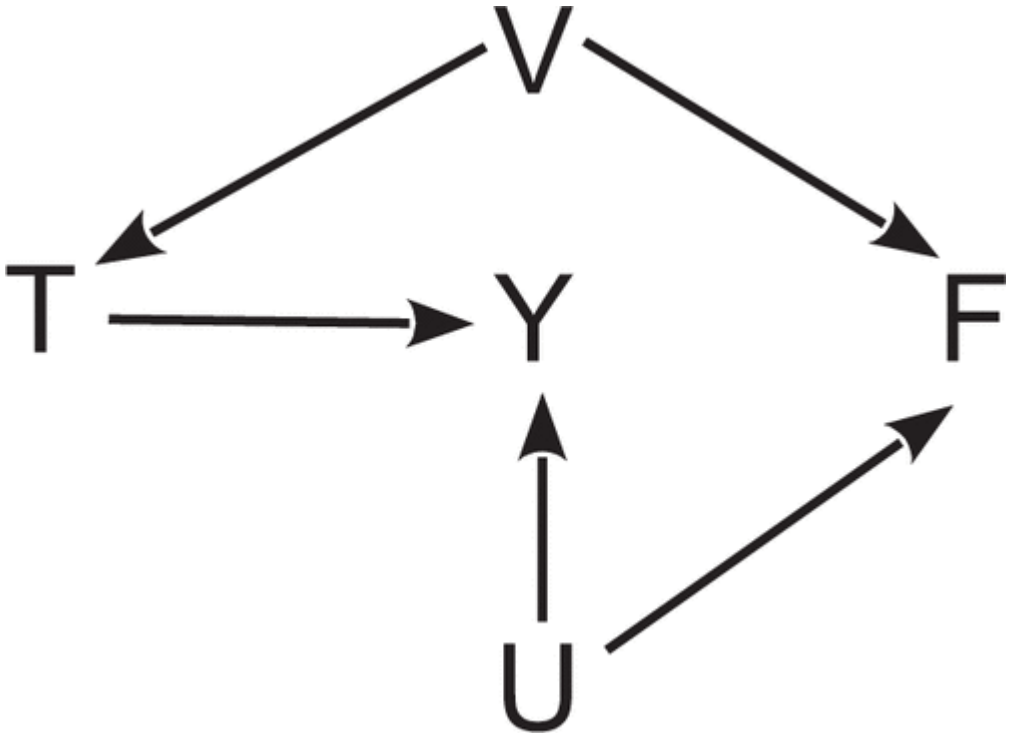

Figure 11. *F* and *Y* are associated even though there is no unobserved confounding between *T* and *Y*. Assumption 1 holds because $T \leftarrow V \rightarrow F$ is always open. *Y* and *F* are associated because $Y \leftarrow U \rightarrow F$ is always open. *T* and *Y* are unconfounded because *F* is an unconditioned collider on the only noncausal path between them, $T \leftarrow V \rightarrow F \leftarrow U \rightarrow Y$.

The asymmetry in result 11 is fixed, and a proper test for unobserved confounding is provided, by assumptions 2 and 3 in result 12.

Definition 5: Let ***Q*** be the nonempty set of unobserved causes of *F*, excluding the idiosyncratic causes of *F* (i.e., excluding the independent error term on *F*).

Assumption 2: All variables in ***Q*** directly cause *T*.

Like assumption 1, assumption 2 is nonparametric, that is, does not assume linearity or effect homogeneity. Nonetheless, assumption 2 is stronger than assumption 1 (which it implies). Assumption 2 requires that all unobserved causes of *F* (except those in *F*s idiosyncratic error term) also cause *T*. This rules out unobserved confounding between *Y* and *F* by any event that occurs after *T*. Assumption 2, however, still does not require that all unobserved confounders of *T* and *Y* also cause *F*, as we had assumed in prior sections of this article (e.g., Figure 2).

Assumption 3: *Y* does not directly or indirectly cause *F* (no selection).

Result 12: Given assumptions 2 and 3, independence between *F* and *Y* conditional on *T* and ***X*** (which may be empty), $F \perp Y|(T, \boldsymbol{X})$, implies the absence of unobserved confounding between *T* and *Y* conditional on ***X***; and nonindependence between *F*

Y conditional on *T* and ***X*** (which may be empty), $F \neg \perp Y | (T, \boldsymbol{X})$, implies the presence of unobserved confounding between *T* and *Y* conditional on **X**. (Proof in Online Appendix C)

Result 12 says that any (nonparametric or parametric) test of conditional independence between *F* and *Y* given *T* and ***X*** is a valid test of nonconfounding between *T* and *Y* given ***X*** when assumptions 2 and 3 are met. Rejection of the conditional independence is evidence of unobserved confounding; and failure to reject conditional independence indicates the absence of unobserved confounding. Translated to the familiar world of linear regression, result 12 says that testing the null hypothesis $H_0: b_{YF.TX} = 0$, against the alternative $H_A: b_{YF.TX} \neq 0$, in the following regression specification,

$$Y_i = a + Tb_{YT.FX} + Fb_{YF.TX} + Xb_{YX.TF} + u_i,$$

using a conventional two-sided *t* test is a valid test of the null hypothesis of no unobserved confounding.[15]

Empirical Illustration

Motivation

We illustrate the utility of future-treatment strategies by elaborating on Mayer's (1997) original empirical example of the causal effect of parental income on children's educational attainment. Parental income is widely observed to be positively associated with children's test scores. This association could be at least partially causal because high income allows parents to invest more in their children's education (Kornrich and Furstenberg 2012; Schneider, Hastings, and LaBriola 2018), for instance, by providing private tutors (Buchmann, Condron, and Roscigno 2010), which promotes educational success (see also Mayer 1997:45ff). On the other hand, the observed association could also be due to unobserved confounding, for example, by parents' ambition, which may increase not only parents' own income but also the educational success of their children. Of more than merely historical interest, this example remains salient for contemporary debates on intergenerational transmission, which are plagued by concerns about unobserved confounding (Morgan and Winship 2015; Sobel 1998).

Data

We analyze data from the Panel Study of Income Dynamics PSID (2019). In an effort to replicate the estimates provided by Mayer (1997:161ff), we closely follow her decisions in

the construction of the analytic samples (covering birth cohorts 1954 through 1968) and variables as described there. The outcome, *Y*, is children's years of education completed by age 24 ($\text{mean} = 12.9$, $SD = 2.0$). The treatment variable, *T*, is logged family income in 1992 dollars, measured at children's ages 13–17 (5-year average). The future-treatment variable, *F*, is logged family income in 1992 dollars, measured at children's ages 25–29 (5-year average). The list of observed confounders, *X*, includes logged family size, whether the child's household head is black, age of the younger parent, the highest years of education attained by either parent, and child's gender. The analytic sample size for the future-treatment regressions is *N* = 1,513. All analyses are weighted by the child's individual survey weight in 1989. Descriptive statistics are given in Online Appendix Table D1. All variables are standardized (mean 0 and variance 1). A replication package for this analysis is available online (http://doi.org/10.3886/E104060V1).

Results

Our analyses replicate Mayer's published results almost perfectly. For instance, Mayer's main analysis (based on cohorts born between 1954 and 1968) estimates the unstandardized coefficient of logged family income on children's years of education as $.78$ $(SE = .07)$, compared to our estimate of $.76$ $(SE = .09)$. We observe an even closer correspondence in our analytic subsample (for which future income measures are available; birth cohorts 1954–1964) with a standardized coefficient estimate of $.19$ in both Mayer's and our own analysis (for full results, see Online Appendix Table D2).

Our empirical illustration of future-treatment strategies estimates a series of OLS models that regress the outcome (*Y*, years of education) on different combinations of the covariates: The treatment (*T*, parental income), the future treatment (*F*, future parental income), and all observed pretreatment control variables mentioned above (X) . All models are estimated from the same analytic data set. Table 3 reports four different model specifications that illustrate the use of different future-treatment strategies under various assumptions about the DGP. Model 1 displays the unadjusted association between parental income (*T*) and offspring's educational attainment (*Y*). Without controlling for any observed confounders (*X*), we expect the association between *T* and *Y* to provide a biased estimate of the causal effect. For illustration purposes, it is helpful to first apply future-treatment strategies to a treatment-effect estimate that we know strongly suspect to be biased. Model 2 adds the future treatment (*F*), to model 1, but does not add any other covariates. The comparison between models 1 and 2 therefore shows how future-treatment strategies produce expected answers in a situation of bias. The more realistic scenario encountered in empirical applications, of course, is that the analyst has alre

attempted to exhaustively adjust for observable differences, reflected in model 3, which adds all control variables used in Mayer's original analyses to model 1. In model 4, we then add a control for the future treatment to model 3 to illustrate the conclusions drawn from future-treatment strategies in the typical empirical setting without prior knowledge about the existence and direction of the remaining bias. We contrast the conclusions drawn based on the control strategy, the Mayer/difference strategy, and our nonparametric test for hidden bias under various assumptions about the DGPs.

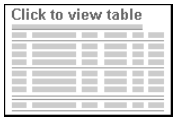

Table 3. Estimating the Causal Effect of Parental Income on Children's Years of Education With and Without Future Treatments.

Best-case Scenario

The best-case scenario for future-treatment estimation is given by the DGP in Figure 2. We recall that this scenario assumes that all confounders of *T* and *Y* also affect *F* and that all confounders of *F* and *Y* also affect *T*. It also assumes the absence of true state dependence (no arrow $T \rightarrow Y$)—that is, changes in parental income during middle childhood (aged 13–17) have no causal impact on parental income during offspring's young adulthood (aged 24–29). Furthermore, it assumes the absence of selection (no arrow $Y \rightarrow F$)—that is, children's years of education do not cause changes in their parents' income.

We begin by testing for the absence of unobserved bias in the unadjusted association between parental income, *T*, and child's educational attainment, *Y*. Model 1 gives this unadjusted association as $b_{YT} = 0.448\ (p < .001)$. The test for the absence of unobserved confounding is implemented by testing the null hypothesis that the *T*-adjusted regression coefficient on *F* is zero, $b_{YF.T} = 0$. In model 2, this null hypothesis is safely rejected ($p < .001$). Hence, we conclude that the naive, unadjusted estimate of model 1 suffers from unobserved bias, which is plausible.

Now that we have provided evidence for in the existence of bias, we use the control strategy to reduce it. The control strategy focuses on the *F*-adjusted coefficient in model 2, $b_{YT.F} = .319\ (p < .001)$, which is significantly smaller than the unadjusted association between parental income and child's educational attainment in model 1 ($b_{YT} - b_{YT.F} = .448 - .319 = .129;\ p < .001$). By result 2, we know that the control strategy is strictly bias reducing under the DGP of Figure 2. Thus, we conclude that the *F*-adjusted estimate from model 2 is closer to the true treatment effect than the naive estimate without

F-adjustment of model 1; the naive treatment effect estimated in model 1 is upwardly biased.

The Mayer/difference method, applied to model 2, estimates the treatment effect as the difference between the partial coefficients on *T* and *F*, $b_{YT.F} - b_{YF.T} = .319 - .274 = .045$ $(p = .723)$. This estimate, too, is lower than the naive estimate of the treatment effect in model 1 $(b_{YT} = .448)$ and also lower than the control estimate $(b_{YT.F} = .319)$. Earlier, we showed that the Mayer/difference estimate is potentially more powerful in reducing bias than the control strategy, but that—unlike the control strategy—it may also amplify bias.

From the application of the control strategy, we learned that the naive estimate of model 1 is upwardly biased. If the Mayer/difference estimator had yielded a larger estimated treatment effect than the naive estimate (cf. Figure 4, scenario 1), we would have concluded that the Mayer/difference strategy amplifies rather than reduces existing bias. If, by contrast, the Mayer/difference estimator had fallen between the naive and the *F*-adjusted estimate of the treatment effect, then, by the argument of Figure 4 (scenario 2), we would have concluded that the difference strategy is less effective in reducing bias than the control strategy. In both instances, we would have preferred the control estimate to the Mayer/difference estimate.

In our application, however, the Mayer/difference estimate moves the naive estimate in the same direction as, but more strongly than, the control estimate (Figure 4, scenario 3). Yet, without further assumptions about the strength of the path parameters, we do not know whether the Mayer/difference estimate is closer to the true causal effect than the control estimate. The most conservative analyst may therefore prefer the estimate provided by the control strategy in this empirical application, noting, however, that bias reduction may be relatively modest unless the effect $U \rightarrow F$ is very large.

In some applications, the analyst may have reasonable expectations about the direction and sign of the effects $U \rightarrow T$, *a* and $U \rightarrow F$, *d*. In our example, an analyst may assume that parents' unobserved ambition, *U*, impacts parental income in the same direction at *T* and *F*, that is, *a* and *d* have the same sign. Then, unless the effect $U \rightarrow F$, *d* is much larger than the effect $U \rightarrow T$, *a*, the analyst should prefer the Mayer/difference estimator as the strategy to reduce the most bias. In sum, in this application, the decision between the control and the difference estimator depends on how defensible the analyst's additional assumptions about the relative size of the effects $U \rightarrow T$ and $U \rightarrow F$ are. If the analyst prefers the Mayer/difference strategy estimates, then one should note that this estimate is not statistically different from zero $(p = .445)$. This would cast doubt on tl

proposition that an increase in parental income causes an improvement in children's educational attainment.

Next, we estimate a treatment effect by controlling for observables (model 3). This covariate-adjusted model estimates the treatment effect as $b_{YT.X} = .185\ (p < .001)$. This estimate is much smaller $(p < .001)$ than the unadjusted association of model 1, which indicates that the control-strategy estimate of model 2 correctly determined the positive direction of the bias in the unadjusted analysis of model 1.

Although, in model 3, we control for a number of important control variables, the careful analyst will still worry about unobserved bias. This worry is addressed in the future-adjusted model 4. Again, the null hypothesis of no unobserved bias cannot be rejected since the coefficient on *F*, $b_{YF.TX} = .202$ is significantly different from zero $(p < .01)$. The presence of unobserved bias motivates future-treatment adjustments in order to reduce bias.

The control estimate of model 4 is smaller than the baseline treatment effect of model 3 ($b_{YT.FX} = .118$ vs. $b_{YT.X} = .185; p < .001$), indicating that controlling for the future treatment has reduced bias. The correction is modest in size but statistically significant $(b_{YT.X} - b_{YT.FX} = .185 - .118 = .067;\ p < .001)$. (The correction was greater going from model 1 to model 2—about twice the size—since the absence of any other control variables in these models put a greater burden of bias reduction on the future treatment.) The Mayer/difference estimate for the treatment effect is yet smaller, and even negative, at $b_{YT.FX} - b_{YF.TX} = -.084\ (p = .098)$, and statistically indistinguishable from zero at conventional levels of statistical significance.[16] As before, absent additional assumptions about the relative strength of the effects $U \rightarrow T$ and $U \rightarrow F$, we cannot be certain that the Mayer/difference estimate is closer to the true treatment effect than the control estimate.

True State Dependence and Selection

When the analyst is not willing to rule out true state dependence and selection, the interpretation of the estimates presented in Table 3 may or may not change. In our empirical example, true state dependence, $T \rightarrow F$, may be present if parents' income growth depends on their baseline income. By result 12, however, even under true state dependence, the test for the absence of unobserved bias remains valid, as long as assumptions 2 and 3 hold. The empirical conclusion of the test would thus remain the same: We would rule out that the treatment-effect estimate is unbiased in all models shown. With state dependence, however, using the control or Mayer/difference estimator to reduce this bias in model 1 requires new assumptions about the size of certain pa

parameters because now even the control estimator is not strictly bias reducing anymore. Most importantly, this includes assumptions about confounding itself. For example, we could assume that the effects $U \rightarrow T$ and $U \rightarrow F$ are of the same size $(a = d)$ and that confounding is of, at most, moderate size $(|a| \leq .5)$. Second, our choice between the control and Mayer/difference estimator is dictated by assumptions about the direction and degree of state dependence. That is, if state dependence is negative or at best weakly positive, the Mayer/difference estimator is the better choice. However, if state dependence is moderately $(f \geq 0.37)$ or strongly positive $(f \geq 0.5)$, the control estimator is the better choice. The stakes involved in making these assumptions are quite high. If they are wrong, future-treatment strategies may amplify bias (namely, the control estimator if state dependence is negative and the difference estimator if state dependence is strongly positive). Existing empirical work on the dynamics of income poverty suggests state dependence to be positive and large (e.g., Biewen 2009; Cappellari and Jenkins 2004), which would lead one to prefer the control estimator.

In our empirical example, selection, $Y \rightarrow F$, would be of concern if, for instance, children's decision to forego college enrollment for entry into the labor force causes parents to reduce their labor supply because they no longer have to pay for children's college tuition. While such selection stories may be plausible for certain subgroups of the population, we are not aware of well-identified estimates of large selection effects. Still, what would selection imply for future-treatment strategies to detect, reduce, and remove bias? Unfortunately, the test for the absence of unobserved bias would no longer be valid under selection (violation of assumption 3). The attractiveness of the difference method is drastically reduced as its bias-amplification property becomes more pronounced. Analysts who believe selection to be a concern in this empirical example should refrain from both an interpretation of the test and of the difference estimator. However, the control estimator would remain useful because it remains bias reducing if confounding is present and selection is mild. Thus, the control estimator would remain the preferred estimate.

Conclusion

The problem of unobserved confounding is profound. Most research in the social sciences is observational and observational studies cannot rule out bias from unobserved confounding. The direction and especially the size of the bias are often difficult to gauge, in part because the bias could originate in confounders that are as yet unknown to science.

In this article, we have discussed future values of the treatment variable as a tool for detecting, reducing, and removing bias from unobserved confounding. Future treatments have occasionally been used for bias removal in prior research. Here, we have subjected several easily computed future-treatment strategies to a detailed analysis, introduced a simple new strategy, and compared the relative strengths and weaknesses of these estimators to each other and to baseline conventional regression estimates. While we identify challenges to future-treatment strategies, we do not stop there. To maximize the usefulness of future-treatment estimators in applied research, we also demonstrate how additional assumptions about effect sizes can help choose between estimators and inform their interpretation.

The idea behind future-treatment strategies is intuitive: Any variable that affects the treatment variable before the outcome likely also affects it after the outcome has been measured. In other words, future treatments can proxy for unobserved confounders. We have used this insight directly and proposed controlling for future treatments as a covariate in a regression (our control estimator). This estimator has the great advantage of being strictly bias reducing for some linear DGPs.

Analyzing important prior future-treatment strategies, we have noted that Mayer's (1997) estimator is not strictly bias reducing even in the best-case scenario and may in fact amplify conventional OLS bias. The same is true of Gottschalk's (1996) future-treatment estimator (Online Appendix A). Nonetheless, Mayer's estimator holds promise because, in certain situations, it reduces bias more than the control estimator.

Future-treatment strategies have several advantages over other strategies for dealing with unobserved confounding. One advantage lies in the ready availability of future-treatment measures in most panel data. Another is the ease of implementation—including future treatments as control variables in a conventional regression analysis. In contrast to fixed-effects estimation, future-treatment strategies to reduce unobserved bias do not require repeated measures of the outcome nor do they require long panels (three periods suffice; see also Vaisey and Miles [2017] for a critical discussion of fixed-effects estimation based on three observation points). Several large social science surveys newly facilitate the application of the future-treatment strategy. For example, recent waves of the General Social Survey included three-wave panels (Hout 2017), and the redesigned Survey of Income and Program Participation includes four-wave panels.

Finally, future-treatment strategies can be used for the dual purpose of detecting and reducing—sometimes even removing—unobserved confounding. Indeed, we have shown

that future treatments can detect the presence of bias even when they cannot reduce this bias.

A limitation shared with all strategies for reducing and removing hidden bias from unobserved confounding is that causal inference always requires detailed knowledge of the DGP. Within the confines of linear and homogenous models, we have highlighted two conditions that pose particular challenges for future-treatment estimators: true state dependence (when prior treatment causes future treatment) and selection (when the outcome causes the future treatment). In both scenarios, all future-treatment estimators may increase or decrease bias in unadjusted OLS estimates. And whereas selection may be ruled out in many substantive applications, true state dependence often remains a credible threat. Based on our analytic results, however, we have argued that the control estimator remains bias reducing for moderate confounding under moderate true state dependence and is surprisingly robust to selection as well.

Since future-treatment strategies make different demands on the DGP than fixed effects or instrumental variables estimators (see Online Appendix B), and because future-treatment measures are widely available in panel data, future-treatment strategies promise help where other popular strategies may fail.

Authors' Note

A replication package containing the data and code used for the empirical illustration in this article is available through the PSID Public Data Extract Repository at https://www.openicpsr.org/openicpsr/psid (#104060).

Acknowledgment

We thank Yongnam Kim and Zeyu Wei for valuable advice and N. E. Barr for copy editing. This work was supported by a grant from the University of Wisconsin Graduate School Research Competition and a Vilas Mid-Career Fellowship from the University of Wisconsin. We gratefully acknowledge use of the services and facilities of the Center for Demography and Ecology at the University of Wisconsin–Madison, funded by NICHD Center Grant P2C HD047873, and the Population Studies Center at the University of Michigan, funded by NICHD Center Grant R24 HD041028. The collection of data used in this study was partly supported by the National Institutes of Health under grant number R01 HD069609 and the National Science Foundation under award number 1157698.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research,

authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by a grant from the University of Wisconsin Graduate School Research Competition and a Vilas Mid-Career Fellowship from the University of Wisconsin; Center for Demography and Ecology at the University of Wisconsin–Madison, funded by NICHD Center Grant P2C HD047873; and the Population Studies Center at the University of Michigan, funded by NICHD Center Grant R24 HD041028. The collection of data used in this study was partly supported by the National Institutes of Health under grant number R01 HD069609 and the National Science Foundation under award number 1157698.

ORCID iD

Fabian T. Pfeffer https://orcid.org/0000-0002-6196-0617

Supplementary Material

Supplementary material for this article is available online.

Notes

1. Other examples of research that purposefully subverts the common temporal order include the correlated random effects model proposed by Chamberlain (1982) as well as applied contributions that consider a comparison group that only experiences treatment in the future, such as future incarceration (e.g., Grogger 1995; Porter and King 2014; Wildeman 2010) or a future network tie (e.g., Kim, Kogut, and Yang 2015).

2. Throughout, we assume large samples in order to focus on identification.

3. Path parameters are often called “path coefficients.” We write “parameter” to denote true causal effects in the DGP, and we write “coefficient” to denote statistical quantities, such as regression coefficients, which may or may not equal the desired parameter.

4. Contrary to conventional wisdom, however, standardized regression coefficients in multivariate models can exceed 1, for example, when multicollinearity is high (Jöreskog 1999).

5. Our formulas apply directly after Y, T, and F have been residualized for other covariates. One typically assumes that controlling for pretreatment variables reduces bias

from unobserved confounding. For counterexamples, see Elwert and Winship (2014) and Steiner and Kim (2016).

6. Additional assumptions embedded in Figure 2 include that (a) all confounders of *F* and *Y* affect *T*, (b) *T* does not cause *F*, (c) *Y* does not cause *F* and (d) *T* and *F* share no unobserved common causes that do not also cause *Y*. We will relax some of these assumptions below.

7. We pick $a = .4$ for illustration. Results are qualitatively the same for other values of *a*.

8. The first two equations are collinear if past and future values of the treatment are very similar, that is, $a = d$ approaches 1. As *a* increases, the denominator of Mayer's estimator, $1 - a^2$, shrinks toward zero. Consequently, standard errors will increase with the magnitude of *a*.

9. When $a \neq d$, the three observable covariances between *T*, *Y*, and *U*, produce three equations with four unknowns: (1) $\sigma_{YT} = b + ac$, (2) $\sigma_{YF} = abd + cd$, and (3) $\sigma_{TF} = ad$, which cannot be solved uniquely for *b*. Mayer (1997:178) proposes an empirical adjustment.

10. True state dependence is also a central challenge in the literature on dynamic treatment effects (Robins 1994; Wodtke and Almirall 2017).

11. Negative state dependence can occur when treatment depletes some fixed stock. For example, if insurance pays only for a limited number of therapy visits, then increasing early visits, *T*, may decrease later visits, *F*. We thank an anonymous reviewer for this suggestion.

12. Judging the equality of $a = d$ in Figure 2 is easier because the structure of the DGP provides that equality of the standardized path parameters corresponds to equality of the parameters on their natural scale.

13. Figure 7 presents an example of postoutcome endogenous selection bias (Elwert and Winship 2014). *Y* is a collider variable on the path $T \rightarrow Y \leftarrow U$, and *F* is a descendant of *Y*. Conditioning on a descendant of a collider induces an association between the collider's immediate causes, that is, between *T* and *U*. Hence, conditioning on *F* induces a noncausal association between *T* and *Y* via *U*, which is the bias in the *F*-adjusted analysis.

14. When the control estimator increases bias, it does so negligibly.

15. Mayer (1997) informally suggested an endogeneity test involving future treatments for the linear DGP of Figure 2. Here, we generalize the test nonparametrically beyond linear models and causally beyond the DGP of Figure 2.

16. While we successfully replicate Mayer's main results, our result using the Mayer/difference estimator is quite different. Mayer's analyses suggest a quite modest drop from 0.186 in model 3 to 0.168 in model 4 Mayer 1997. Ours show a much larger drop from 0.185 to a statistically insignificant estimate of −0.084. The conclusions that may be drawn from our estimate—no evidence for a causal relationship between parental income and children's educational attainment—are in fact more supportive of Mayer's general conclusions. http://proceedings.mlr.press/v9

References

Bates, G. E., Neyman, J.. 1951. "Contributions to the Theory of Accident Proneness. An Optimistic Model of the Correlation between Light and Severe Accidents." University of California Publications in Statistics 1:215–54.
Google Scholar

Biewen, Martin . 2009. "Measuring State Dependence in Individual Poverty Histories When There Is Feedback to Employment Status and Household Composition." Journal of Applied Econometrics 24:1095–116.
Google Scholar | Crossref

Brito, Carlos, Pearl, Judea. 2002. "Generalized Instrumental Variables." Pp. 85–93 in Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence, edited by Darwiche, A., Friedman, N.. San Francisco, CA: Morgan Kaufmann.
Google Scholar

Buchmann, Claudia, Condron, Dennis, Roscigno, Vincent. 2010. "Shadow Education, American Style. Test Preparation, the SAT and College Enrollment." Social Forces 89:435–62.
Google Scholar | Crossref | ISI

Cappellari, Lorenzo, Jenkins, Stephen P.. 2004. "Modelling Low Income Transitions." Journal of Applied Econometrics 19:593–610.
Google Scholar | Crossref

Chamberlain, Gary . 1982. "Multivariate Regression Models for Panel Data." Journal of Econometrics 18:5–46.
Google Scholar | Crossref | ISI

Chan, Hei, Kuroki, Manabu. 2010. "Using Descendants as Instrumental Variables for the Identification of Direct Causal Effects in Linear SEMs." Pp. 73–80 in International Conference on Artificial Intelligence and Statistics.
Google Scholar

Deluca, Stefanie . 2012. “What Is the Role of Housing Policy? Considering Choice and Social Science Evidence.” Journal of Urban Affairs 34:21–8.
Google Scholar | Crossref | Medline | ISI

DiNardo, John, Pischke, Jorn-Steffen. 1997. “The Returns to Computer Use Revisited: Have Pencils Changed the Wage Structure Too?” Quarterly Journal of Economics 112:291–303.
Google Scholar | Crossref | ISI

DiPrete, Thomas A., Eirich, Gregory M.. 2006. “Cumulative Advantage as a Mechanism for Inequality: A Review of Theoretical and Empirical Developments.” Annual Review of Sociology 32:271–97.
Google Scholar | Crossref | ISI

Duncan, Greg J . 2017. “Household Income and Child Development in the First Three Years of Life.” NIH Grant 1R01HD087384-01A1.
Google Scholar

Duncan, Greg J., Connell, James P., Klebanov, Pamela K.. 1997. “Conceptual and Methodological Issues in Estimating Causal Effects of Neighborhoods and Family Conditions on Individual Development.” Pp. 219–50 in Neighborhood Poverty, Volume 1: Context and Consequences for Children, edited by Brooks-Gunn, Jeanne, Duncan, Greg J., Lawrence Aber, J.. New York, NY: Russell Sage.
Google Scholar

Elwert, Felix . 2013. “Graphical Causal Models.” Pp. 245–73 in Handbook of Causal Analysis for Social Research, edited by Morgan, S. L. . Dordrecht, the Netherlands: Springer.
Google Scholar | Crossref

Elwert, Felix, Christakis, Nicholas A.. 2008. “Wives and Ex-wives: A New Test for Homogamy Bias in the Widowhood Effect.” Demography 45:851–73.
Google Scholar | Crossref | Medline | ISI

Elwert, Felix, Winship, Christopher. 2014. “Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable.” Annual Review of Sociology 40:31–53.
Google Scholar | Crossref | Medline | ISI

Gottschalk, Peter . 1996. “Is the Correlation in Welfare Participation across Generations Spurious?” Journal of Public Economics 63:1–25.
Google Scholar | Crossref | ISI

Grogger, J. 1995. “The Effect of Arrests on the Employment and Earnings of Young Men.” The Quarterly Journal of Economics 110:51–71.
Google Scholar | Crossref | ISI

Heckman, James J. 1981a. “Heterogeneity and State Dependence.” Pp. 91–140 in Studies in Labor Markets, edited by Rosen, S. . Chicago, IL: University of Chicago Press.
Google Scholar

Heckman, James J . 1981b. “Statistical Models for Discrete Panel Data.” Pp. 114–78 in Structural

Analysis of Discrete Data and Econometric Applications, edited by Manski, Charles F., McFadden, Daniel L.. Cambridge: MIT Press.
Google Scholar

Hout, Michael . 2017. “Models for Three-wave Panel Data: Examples Using the General Social Survey Panels.” Sociological Methods & Research 46:41–3.
Google Scholar | SAGE Journals | ISI

Jöreskog, Karl G. 1999. “How Large Can Standardized Coefficients Be?” Unpublished Manuscript. Retrieved April 16, 2019 (https://wenku.baidu.com/view/f5159e1d10a6f524cdbf8500.html).
Google Scholar

Kim, Jerry W., Kogut, Bruce, Yang, Jae-Suk. 2015. “Executive Compensation, Fat Cats, and Best Athletes.” American Sociological Review 80:299–328.
Google Scholar | SAGE Journals | ISI

Kim, Yongnam, Steiner, Peter. in press. “Gain Scores Revisited: A Graphical Models Perspective.” Sociological Methods & Research. https://journals.sagepub.com/doi/full/10.1177/0049124119826155.
Google Scholar

Kornrich, Sabino, Furstenberg, Frank. 2012. “Investing in Children. Changes in Parental Spending on Children, 1972–2007.” Demography 50:1–23.
Google Scholar | Crossref

Mayer, Susan E. 1997. What Money Can’t Buy. Family Income and Children’s Life Chances. Cambridge, MA: Harvard University Press.
Google Scholar

Morgan, Stephen L., Winship, Christopher. 2015. Counterfactuals and Causal Inference. Methods and Principles for Social Research. 2nd ed. Cambridge, MA: Cambridge University Press.
Google Scholar

Pearl, Judea . 2009. Causality: Models, Reasoning, and Inference. 2nd ed. Cambridge, MA: Cambridge University Press.
Google Scholar | Crossref

Pearl, Judea . 2013. “Linear Models: A Useful ‘Microscope’ for Causal Analysis.” Journal of Causal Inference 1:155–70.
Google Scholar | Crossref

Panel Study of Income Dynamics (PSID) . 2019. Panel Study of Income Dynamics, Public Use Dataset. Produced and Distributed by the Survey Research Center, Institute for Social Research. Ann Arbor, MI: University of Michigan.
Google Scholar

Piketty, Thomas . 2014. Capital in the Twenty-first Century. Cambridge, England: Belknap Press.
Google Scholar | Crossref

Porter, Lauren C., King, Ryan D.. 2014. “Absent Fathers or Absent Variables? A New Look at Paternal Incarceration and Delinquency.” Journal of Research in Crime and Delinquency 52:414–43.
Google Scholar | SAGE Journals

Robins, James M. 1994. “Correcting for Non-compliance in Randomized Trials Using Structural Nested Mean Models.” Communications in Statistics—Theory and Methods 23:2379–412.
Google Scholar | Crossref | ISI

Rosenbaum, Paul R. 2002. Observational Studies. 2nd ed. New York, NY: Springer.
Google Scholar | Crossref

Schneider, Daniel, Hastings, Orestes P., LaBriola, Joe. 2018. “Income Inequality and Class Divides in Parental Investments.” American Sociological Review 83:475–507.
Google Scholar | SAGE Journals | ISI

Steiner, Peter, Kim, Yongnam. 2016. “The Mechanics of Omitted Variable Bias: Bias Amplification and Cancellation of Offsetting Biases.” Journal of Causal Inference 4:983982.
Google Scholar

Sobel, Michael E. 1998. “Causal Inference in Statistical Models of the Process of Socioeconomic Achievement.” Sociological Methods & Research 27:318–48.
Google Scholar | SAGE Journals | ISI

Vaisey, Stephen, Miles, Andrew. 2017. “What You Can—and Can’t—Do with Three-wave Panel Data.” Sociological Methods & Research 46:44–67.
Google Scholar | SAGE Journals | ISI

Verma, Tom S., Pearl, Judea. 1988. “Causal Networks: Semantics and Expressiveness.” Proceedings of the Fourth Workshop on Uncertainty in Artificial Intelligence. https://www.sciencedirect.com/science/article/pii/B9780444886507500111
Google Scholar

Wildeman, Christopher . 2010. “Paternal Incarceration and Children’s Physically Aggressive Behaviors: Evidence from the Fragile Families and Child Wellbeing Study.” Social Forces 89:285–309.
Google Scholar | Crossref | ISI

Wodtke, Geoffrey T., Almirall, Daniel. 2017. $U \rightarrow F$ “Estimating Moderated Causal Effects with Time-varying Treatments and Time-varying Moderators: Structural Nested Mean Models and Regression with Residuals.” Sociological Methodology 47:212–45.
Google Scholar | SAGE Journals | ISI

Wright, Sewall . 1921. “Correlation and Causation.” Journal of Agricultural Research 20:557–85.
Google Scholar

Author Biographies

Felix Elwert is Romnes Professor of Sociology and of Biostatistics and Medical Informatics at the University of Wisconsin-Madison. His research concerns methods

causal inference, social stratification, and spillover processes in social networks.

Fabian T. Pfeffer is an associate professor in the Department of Sociology and Associate Research Professor at the Institute for Social Research at the University of Michigan. He is also a Co-Investigator of the Panel Study of Income Dynamics and Founding Director of the Center for Inequality Dynamics. His research investigates social inequality and its maintenance across time and generations.

Similar Articles:

View all >

Gain Scores Revisited: A Graphical Models Perspective

Yongnam Kim and more...

Sociological Methods & Research

Feb 2019

Regression-based Adjustment for Time-varying Confounders

Geoffrey T. Wodtke

Sociological Methods & Research

May 2018

Interactions in Fixed Effects Regression Models

Marco Giesselmann and more...

Sociological Methods & Research

Apr 2020