Back to the Future: Controlling for Future Treatments to Assess Hidden Bias

Felix Elwert

University of Wisconsin-Madison

Fabian T. Pfeffer

University of Michigan

— Draft, not for circulation —

Abstract

Hidden bias from unobserved confounding is a central problem of causal inference from observational data. One strategy for mitigating hidden bias previously employed in population sciences is to control for *future* (i.e. post-outcome) values of the treatment. The basic idea is that the unobserved confounders affecting treatment likely also affect future values of the treatment. If so, future values of the treatment can proxy for the unmeasured confounder, and controlling for the proxy may remove part of the bias.

This paper investigates the utility of future treatments to control for hidden bias. Drawing on the theory of directed acyclic graphs (Pearl 1995, 2009) we state the nonparametric conditions under which this strategy succeeds in reducing bias, and when it does not. We also state some parametric considerations and explain how future treatments can be used to detect the direction of bias. Centrally, we sketch a new parametric test for the presence of unobserved confounding. We illustrate these results with an empirical example, namely the estimation of the effects of parental wealth on children's probability of graduating from high school.

1 Introduction

Hidden bias from unobserved confounding is a central problem of causal inference from observational data (e.g., Rosenbaum 2002, Elwert and Christakis 2008). One popular strategy for mitigating hidden bias previously employed in sociology and economics is to control for *future* (i.e. post-outcome) values of the treatment. The basic idea is that the unobserved confounders affecting treatment likely also affect future values of the treatment. If so, future values of the treatment can proxy for the unmeasured confounder, and controlling for the proxy may remove part of the bias.

For example, Mayer (1997) investigates the causal effect of current family income on various child outcomes, including educational achievement and attainment, teen pregnancy, and single motherhood. She controls for future income to capture potential confounders, such as parental practices or cognitive skills, and interprets the resulting decreased point-estimates of the income effects as evidence for the non-causal role of current income on child outcomes. Greg Duncan and collaborators (1997) apply a future treatment strategy to lend credence to the causal role of neighborhood characteristics on children's IQ. They assume that controlling for selection into future neighborhoods will control for selection into current neighborhoods. Further examples from labor economics range from an assessment of union wage effects (Chamberlain 1982) to the wage effects of smoking (Stafford and Grafova 2009). Controlling for future union status or smoking status is supposed to elucidate whether these factors play a causal role in determining wage levels.

This paper investigates the utility of future treatments to control for hidden bias. First, drawing on Pearl's (1995, 2009) theory of directed acyclic graphs (DAGs) we state the nonparametric conditions under which this strategy succeeds in reducing bias, and when it does not. Second, we state some parametric considerations and

explain when future treatments can be used to bound the true causal effect. Third, we sketch a parametric test for the presence of unobserved confounding. Finally, we illustrate these methodological considerations with an empirical example, namely the estimation of the effects of parental wealth on children's probability of graduating from high school (Pfeffer 2010).

2 The Problem

Figure 1 strips the underlying confounding challenge down to the basics. Consider an

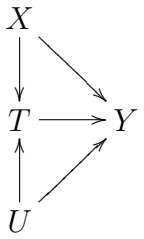

Figure 1

observational study to estimate the causal effect of some treatment, T (e.g. parental wealth), on some outcome, Y (e.g. high school graduation). Since treatment is not randomized, the effect of T on Y may be confounded by factors that jointly affect treatment and outcome, such that the unadjusted association between T and Y will be biased for the causal effect of T on Y. Typically, some confounding variables, X, are measured. If all confounding variables are measured, then conditioning on X will remove all bias. If, however, there also exist unmeasured confounders, U, then the X-adjusted association between T and Y will remain biased. In the following, we will assume that the analyst has appropriately adjusted for all observed direct causes of T, and hence omit X from the following figures for parsimony.

The problem of unmeasured confounding is profound. In observational studies, the

persistent skeptic who suspects the existence of some lurking unobserved confounder, U, can never be proven wrong – U could represent almost anything, including confounders not yet known to science (Rosenbaum 2002).

3 When Controlling for Future Treatments Works

Controlling for future treatments holds promise because it gets around the problem of having to measure, or even know, the nature of U. Figure 2 describes the ideal case.

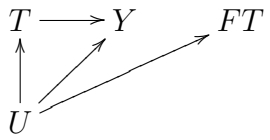

Figure 2

As before, we consider the situation where the causal effect of T on Y is confounded in an unobserved variable U. Even without knowing the nature of U, the analyst may be willing to assume that any U affecting treatment, T, also affects future values of the treatment, FT. If – as in Figure 2 – FT is directly affected only by U, but not by any other variable in the DAG, then conditioning on FT will mitigate the bias in the estimate for the causal effect of T on Y, since FT serves as proxy for the unobserved confounder, U.

FT's ability to proxy for confounding in U depends on the strength of the association between U and FT. If U is perfectly correlated with FT then conditioning on FT is equivalent to conditioning on U itself, and all bias is removed. The smaller the association between U and FT, the weaker FT's ability to proxy for U. Since U and FT will rarely be perfectly correlated, controlling for FT will realistically remove some, but not all, bias.

4 When Controlling for Future Treatments May Not Work

Despite its intuitive appeal, the strategy of controlling for future treatments breaks down if (i) past treatment affects future treatment, or (ii) the outcome affects future treatment, or both. These claims are best demonstrated by showing that controlling for a future treatment can create bias even if the relationship between treatment and outcome is unconfounded.

When Treatment Affects Future Treatment

Past values of the treatment often strongly predict future values of the treatment. This possibility is encoded in Figure 3, where T has a direct effect on FT. For simplic-

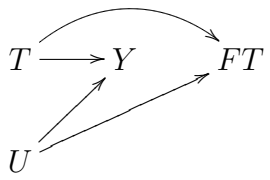

Figure 3

ity, Figure 3 assumes that the effect of T on Y is unconfounded in U (no arrow from U into T) such that estimation could proceed without further adjustment. Needlessly controlling for FT, however, would introduce bias because FT is a "collider" variable (Pearl 2009) on the non-causal path T→FT←U→Y that connects treatment and outcome. Controlling for FT would "unblock" this path, inducing a new, non-causal association between treatment and U, and thus between treatment and outcome. In other words, controlling for FT in Figure 3 would induce an association between treatment, T, and the error term on Y, U. Controlling for FT would therefore create bias where none existed before. Embellishing the DAG in Figure 3 by adding additional

arrows (i.e. relaxing exclusion restrictions) does not change the essential conclusion that controlling for a future outcome that is itself affected by past treatment can bias an otherwise unbiased estimate.

Notice that the existence of the direct effect of T on FT is not testable in Figures 2 and 3. Indeed, the DAGs of Figures 2 and 3 are empirically indistinguishable (absent parametric assumptions) because they imply the same observable qualitative associations. Data alone, in the absence of strong theory, can therefore not exclude the possibility that controlling for FT will increase rather than decrease bias.

If, however, U in Figure 3 were to affect T directly, such that the association between T and Y is already biased for the causal effect without controlling for FT, then the non-causal association induced between T and Y by controlling for FT may increase or decrease this original bias. As we discuss below, with strong additional parametric assumptions, it may occasionally be possible to specify the direction of the bias and use it to the analyst's advantage.

When The Outcome Affects Future Treatment

The strategy of controlling for a future treatment may also fail if the outcome affects the future treatment, as shown in Figure 4. Note that U in Figure 4 is simply

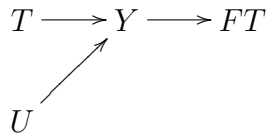

Figure 4

an *a priori* benign idiosyncratic error term on T that does not confound the effect of T on Y. Unbiased estimation of T→Y would thus be possible without further adjustments. Conditioning on FT, however, will introduce bias because T is a collider on the path T→Y←U, and FT is Y's descendant. Conditioning on a descendant of

the outcome is qualitatively the same as selecting on the outcome itself (because FT carries information about Y), which is well known to cause bias by inducing an association between treatment, T, and the error term, U. Controlling for a future treatment affected by the outcome will therefore create bias where none existed before.

As before, embellishing the DAG in Figure 4 by relaxing exclusion restrictions and drawing additional arrows does not change this essential conclusion. Furthermore, although the existence of T→FT is testable in Figure 4, it would no longer be testable if U were to cause FT, which is the very possibility that would motivate controlling for FT in the first place by rendering FT a proxy for U. It is therefore not generally possible to infer from data alone in the absence of strong theory whether controlling for FT would decrease bias (as in Figure 2) or create bias in the first place (as in Figure 4).

5 Some parametric considerations

The direction of the bias induced by controlling for FT is not generally predictable absent strong parametric assumptions. One such set of assumptions, popularly invoked in demography and the social sciences, is embedded in linear path models with continuous variables and iid ~N errors. In such models, we can specify simple conditions under which controlling for future treatment will reduce or exacerbate bias. Consider, for example, the linear path model compatible with Figure 5. The defining exclusion restriction of Figure 5 is that the outcome, Y, does not affect future values of the treatment, FT. This is often plausible in empirical applications. For example although parents' home value, T, likely affects children's high school graduation, Y, high school graduation in turn is quite unlikely to affect parents' future home value, FT, the following year. As before, we suspect that unobservables, U, confound the

causal effect of T on Y. It is trivial to show that the size of the bias in the regression of Y on T depends on the two path coefficients a and d. Further controlling for FT in the regression, by the arguments presented above, will contribute a non-causal association to the estimate, which depends on the path coefficients b, c, and d.

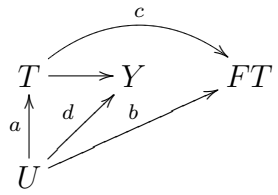

Figure 5

A number of useful results can be derived from these facts. Specifically, controlling for FT in the regression of Y on T in data compatible with Figure 5 will

1. Eliminate all bias if the bias due to U and the bias induced by controlling for FT cancel each other out exactly, which occurs when $a{=}bc$;

2. Reduce bias if $a{<}bc{<}0$ or $0{<}bc{<}a$;

3. Increase bias if $a{<}0{<}bc$ or $bc{<}0{<}a$; and

4. Revert the sign of the bias if $bc{<}a{<}0$ or $0{<}a{<}bc$.

Further consideration of empirical regularities gives additional purchase. Since it is often plausible to assume that $c{>}0$ (because past treatment predicts future treatment) and that a and b have the same sign (because the unobserved confounder likely affects past and future treatment similarly), a and bc are likely of the same sign, which would rule out an increase in bias.

It may be more hazardous, however, to speculate whether a or bc is closer to 0. If in addition to assuming that a and bc are of the same sign one could be sure that a is closer to 0 than bc, then controlling for FT would change the size of the bias and

also reverse the sign of the bias. Therefore, the two estimates arrived by adjusting for FT, and not adjusting for FT, respectively, would bound the true causal effect.

6 A non-parametric test for hidden bias

We can use future treatments for testing the Null hypothesis of no unobserved confounding in the causal effect of T on Y. Specifically, if we are willing to assume that all unobserved direct causes of T also directly cause FT, then the absence of change in the association between T and Y after controlling for FT indicates the absence of confounding in T→Y. The justification for this test is that all confounding of T→Y must either originate, or be mediated by, unobserved direct causes of T; if all unobserved direct causes of T are also direct causes of FT then FT proxies for these unobserved causes, and conditioning on FT will change the association between T and Y. This is a helpful result because it informs the analyst when worries about the presence of hidden bias are misplaced.

Note, however, that this test offers a sufficient but not a necessary test for absence of hidden bias: it is possible that T→Y is unconfounded even if conditioning on FT changes the association, as previously shown in Figures 3 and 4. Furthermore, if the association between T and Y changes as a result of controlling for FT we cannot generally say whether the adjusted estimate is more or less biased than the unadjusted estimate absent certain structural and parametric assumptions (see sections 3 and 5). Naturally, all caveats regarding statistical testing (power, significance) apply.

Even though this test is extremely conservative—in the sense of possibly detecting hidden bias where none exists—it may be rather useful in practice. First, it suits the conservative data analyst to err on the side of caution. Second, it can be used widely as long as the treatment is repeated and the data are longitudinal, that is for most

sociologically interesting treatments in most panel datasets.

7 Application: Effects of parental wealth on children's educational outcomes

We illustrate these arguments with an empirical example of the effects of parents' wealth on their children's educational outcomes (see e.g., Conley 1999, Conley 2001, Morgan and Kim 2006, Belley and Lochner 2007, Pfeffer 2011). We choose this example because the wide-spread research interest in the intergenerational transmission of status has, in recent years, been accompanied by a lively methodological debate on the very possibility of causal inference in this area (Sobel 1998, Morgan and Kim 2006).

Suppose that an analyst wants to know whether and how the observed association between parental wealth and educational outcomes suffers from unobserved bias for the true causal effect. Standard neo-classical economics and rational choice theory, among others, raise serious reasons for concern. Wealth is viewed as a postponement of consumption. In a world of perfect credit markets, the distribution of wealth would simply follow from differential savings propensities that can arise from a range of different individual characteristics: individuals' discount rates applied to future earnings, risk aversion, and others. If these or other unobserved parental characteristics not only determine the propensity to accumulate assets but also influence the educational outcomes of their children, then estimates for the effects of wealth will be biased.

The data for this exercise come from the Child Development Supplement (CDS) to the Panel Study of Income Dynamics (PSID) (see also Williams Shanks 2007; Ye-

ung and Conley 2008; Elliott et al. 2011). We select children who were between the ages 8 and 12 in 1997, who completed a math achievement test, and whose parents had responded to the PSID main questionnaire in 1994 (N=926). The outcome of interest (Y) is the the child's math achievement in 1997, measured as the broad math score on the Woodcock- Johnson Revised Tests of Achievement ranging from 18 to 184. The treatment of interest is parents' home value as one important component of parental wealth (averaged across the years 1989 and 1994 to reduce measurement error, corrected for outliers, logged)[1]. The future treatment (FT) is parents' home value following the assessment of children's math ability averaged across years 1999 and 2001. The list of observed potential confounders (X) includes parents' permanent income (averaged across eight years, 1989-1996), highest education (years and degree), highest occupational status (SES), and an indicator for parental unemployment; grandparents' highest education; the household head's age, sex, and marital status; the number of children in the household; urbanicity; child's age and race; whether the child was of low birthweight; whether the mother received AFDC while pregnant; mother's cognitive ability; and parental risk tolerance (see also Yeung and Conley 2008). Our analysis focuses on the change in the estimated treatment effect for parents' home value across several OLS regression models that include different combinations of regressors: the treatment, the future treatment, and observed control variables.

[1]In another iteration of this paper before the PAA conference, we will also discuss the effect of other components of parental wealth that illustrate additional aspects of the methodological challenges involved (such as financial assets, for which some key assumptions outlined in section 4 cannot be maintained as convincingly).

Testing for bias

Table 1 applies our "greedy" test for the presence of unobserved bias (section 6) to the effect of home values on math achievement. The comparison between the treatment effect estimates in Models I and II establishes the presence of observable bias in the unadjusted association: the treatment effect estimate decreases upon introducing controls for observed potential confounders. As noted in section 6, a change in the association between treatment and outcome after controlling for the future treatment could indicate unobserved bias in the treatment effect estimate, whereas the absence of a change would indicate the absence of bias. We apply this test by comparing the wealth effect estimate from model II (the standard regression model) to the wealth effect estimate in model III (the future treatment model). The change in the wealth coefficient is not statistically significant. We therefore conclude that there is no statistically significant evidence for bias in the treatment effect estimate of Model II. Note that the analyst's confidence in the conclusion of no unobserved bias depends on validity of the central underlying assumption, that is that all unobserved direct causes of wealth also cause future wealth.

Table 1: Effect of parents' home value on children's math achievement

		I	II	III	IV
T:	Home value [in $10k]	.889	.405	.400	.614
		(.070)	(.106)	(.120)	(.116)
FT:	Future home value [in $10k]			.004	.285
				(.104)	(.095)
X:	Controls	no	yes	yes	no

We can further check on the credibility of the test by testing for the presence of the observed bias that we know to exist. Specifically, we have already established that Model I, which does not include any observed control variables, is biased. Adding future wealth to this model (model IV) leads to a substantial and statistically sig-

nificant change in the estimated treatment effect. The test thus accurately detects a bias that we knew exists. Nevertheless, we reiterate that, while we know that bias is present in this specific example, the test is greedy in the sense that it can also lead to a change in coefficients when no bias exists; the absence of a change, however, does indicate the absence of hidden bias.

Direction of bias

Forget, for the moment, that the above analysis did not detect statistically significant evidence for the presence of bias, and suppose instead that the analyst considers the reported changes in treatment coefficients to be substantial. Our discussion in Sections 3 and 6 could then offer further insights into whether the models including future wealth provide a more or a less biased estimate for the desired treatment effect than the conventional models excluding future wealth.

The analysis of section 3 showed that the future-adjusted model will be less biased (regardless of functional form) if (1) the same unobservables affect home values, the outcome, and future home values; (2) math achievement does not affect future home values; and (3) home values do not affect future home values. This last assumption, at a minimum, is clearly implausible – current home wealth directly affects future home wealth. The analyst should therefore not immediately prefer the future-adjusted estimates.

Under less onerous structural assumptions (but at the price of additional parametric assumptions), section 6 offers additional guidance. The analyst requires two structural assumptions.

1. The analyst has to assume that at least some unobserved confounders of the effect of home values on math achievement also affect future home values. This

seems indeed plausible.

2. The analyst must assume that educational outcomes do not affect future parental wealth. This assumption seems plausible for the example of math achievement and home values, but would be considerably less convincing for other specifications of wealth effects on educational outcomes, such as the effect of parents' financial wealth on their children's college attainment. As many parents of college students may attest, college attainment can indeed drain parental finances. By contrast, the analyst may be much more satisfied that math achievement in primary and middle school does not causally affect parent's home value and thus continue the exercise. (Recall that it is not possible to test whether math scores affects future home values because the analyst has previously assumed that the two are confounded in observed variables.)

Under the foregoing two structural assumptions, the parametric considerations of section 6 apply. Start by supposing that wealth causes future wealth (i.e., that the path coefficient $c > 0$). Then

- Bias would increase if the effect of U on wealth has a different sign than its effect on future wealth.This scenario appears unlikely in this application since we can think of no compelling theory for sign reversal – most unobserved U can be expected to affect current and future wealth in the same direction. If so, then future-adjustment does not increase bias in the treatment effect estimate.

- Bias would decrease if the effects of U on wealth and future wealth are of the same sign, and if the effect of U on wealth it is stronger than the product of the effect of wealth on future wealth times the effect of U on future wealth. As suggested above, the first part of this condition is likely to be satisfied, but we

are aware of no theoretical basis for rendering a judgment about the likelihood of the second condition.

- The degree of bias changes and the sign of the bias would be reversed if the effect of U on wealth is of the same sign and if it is weaker than the product of the effect of wealth on future wealth times the effect of U on future wealth. Again, the veracity of the second part of this assumption is difficult to judge. But this situation would imply that the formerly upwardly biased estimate is now downwardly biased, or vice versa. The true effect would thus lie somewhere in between the baseline estimate and the future-adjusted estimate.

In summary, this discussion in the present empirical context suggests that it appears unlikely that controls for future wealth increase bias in home wealth effects. Instead, adjusting for future wealth either decreases bias, or over-adjusts for bias, effectively telling us that the originally estimated treatment effect is upwardly biased. Since both the conventional estimate and the future-adjusted estimate for the effect of parental home wealth on math achievement are positive, substantively large, and statistically significant – and if the analyst believes the above-mentioned assumptions – then the analyst would be justified in expressing greater confidence in the existence of a causal effect of home wealth on educational achievement (see also Haurin et al. 2002).

References

Belley, Philippe and Lance Lochner. 2007. "The Changing Role of Family Income and Ability in Determining Educational Achievement." *Journal of Human Capital* 1:37–89.

Chamberlain, Gary. 1982. "Multivariate Regression Models for Panel Data." *Journal of Econometrics* pp. 5–46.

Conley, Dalton. 1999. *Being Black, Living in the Red. Race, Wealth, and Social Policy in America*. Berkeley: University of California Press.

Conley, Dalton. 2001. "Capital for College. Parental Assets and Postsecondary Schooling." *Sociology of Education* 74:59–72.

Duncan, Greg J., James P. Connell, and Pamela K. Klebanov. 1997. "Conceptual and Methodological Issues in Estimating Causal Effects of Neighborhoods and Family Conditions on Individual Development." In *Neighborhood Poverty. Context and Consequences for Children*, edited by Jeanne Brooks-Gunn, Greg J. Duncan, and J. Lawrence Aber, pp. 219–250. Russell Sage.

Elliott, William, Hyunzee Jung, Kevin Kim, and Gina Chowa. 2011. "A multi-group structural equation model (SEM) examining asset holding effects on educational attainment by race and gender." *Journal of Children and Poverty* 16:91–121.

Elwert, Felix and Nicholas A. Christakis. 2008. "Wives and Ex-Wives. A New Test for Homogamy Bias in the Widowhood Effect." *Demography* 45:851–873.

Haurin, Donald R., Toby L. Parcel, and R. Jean Haurin. 2002. "Does Homeownership Affect Child Outcomes?" *Real Estate Economics* 30:635–666.

Mayer, Susan E. 1997. *What Money Can't Buy. Family Income and Children's Life Chances*. Cambridge: Harvard University Press.

Morgan, Stephen L. and Young-Mi Kim. 2006. "Inequality of Conditions and Intergenerational Mobility. Changing Patterns of Educational Attainment in the United States." In *Mobility and Inequality. Frontiers of Research in Sociology and Economics*, edited by Stephen L. Morgan, David B. Grusky, and Gary S. Fields, pp. 165–194. Stanford: Stanford University Press.

Pearl, Judea. 1995. "Causal diagrams for empirical research." *Biometrika* 82:669–710.

Pearl, Judea. 2009. *Causality. Models, Reasoning, and Inference*. Cambridge: Cambridge University Press, 2nd edition edition.

Pfeffer, Fabian T. 2010. *Wealth and Opportunity in the United States and Germany*. Dissertation. Madison: University of Wisconsin.

Pfeffer, Fabian T. 2011. "Status Attainment and Wealth in the United States and Germany." In *Persistence, Privilege, and Parenting. The Comparative Study of Intergenerational Mobility*, edited by Timothy M. Smeeding, Robert Erikson, and Markus Jäntti. New York: Russell Sage Foundation.

Rosenbaum, Paul R. 2002. *Observational Studies*. New York: Springer, 2nd edition edition.

Sobel, Michael E. 1998. "Causal Inference in Statistical Models of the Process of Socioeconomic Achievement." *Sociological Methods & Research* 27:318–348.

Stafford, Frank and Irina Grafova. 2009. "Wage Effects of Personal Smoking History." *Industrial and Labor Relations Review* 26:381–393.

Williams Shanks, Trina R. 2007. "The Impacts of Household Wealth on Child Development." *Journal of Poverty* 11:93–116.

Yeung, W. Jean and Dalton Conley. 2008. "Black-White Achievement Gap and Family Wealth." *Child Development* 79:303–324.