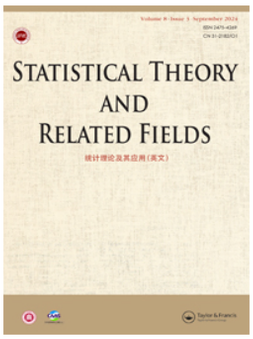


Statistical Theory and Related Fields

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tstf20

A confounding bridge approach for double negative control inference on causal effects

Wang Miao, Xu Shi, Yilin Li & Eric J. Tchetgen Tchetgen

To cite this article: Wang Miao, Xu Shi, Yilin Li & Eric J. Tchetgen Tchetgen (30 Aug 2024): A confounding bridge approach for double negative control inference on causal effects, Statistical Theory and Related Fields, DOI: 10.1080/24754269.2024.2390748

To link to this article: https://doi.org/10.1080/24754269.2024.2390748

STATISTICAL THEORY AND RELATED FIELDS
https://doi.org/10.1080/24754269.2024.2390748

A confounding bridge approach for double negative control inference on causal effects

Wang Miao[a], Xu Shi[b], Yilin Li[a] and Eric J. Tchetgen Tchetgen[c]

[a]Department of Probability and Statistics, Peking University, Beijing, People's Republic of China; [b]Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA; [c]Department of Statistics and Data Science, University of Pennsylvania, Philadelphia, PA, USA

ABSTRACT

Unmeasured confounding is a key challenge for causal inference. In this paper, we establish a framework for unmeasured confounding adjustment with negative control variables. A negative control outcome is associated with the confounder but not causally affected by the exposure in view, and a negative control exposure is correlated with the primary exposure or the confounder but does not causally affect the outcome of interest. We introduce an outcome confounding bridge function that depicts the relationship between the confounding effects on the primary outcome and the negative control outcome, and we incorporate a negative control exposure to identify the bridge function and the average causal effect. We also consider the extension to the positive control setting by allowing for the nonzero causal effect of the primary exposure on the control outcome. We illustrate our approach with simulations and apply it to a study about the short-term effect of air pollution on mortality. Although a standard analysis shows a significant acute effect of PM2.5 on mortality, our analysis indicates that this effect may be confounded, and after double negative control adjustment, the effect is attenuated toward zero.

ARTICLE HISTORY
Received 25 October 2023
Revised 12 April 2024
Accepted 31 July 2024

KEYWORDS
Air pollution effect; confounding; instrumental variable; negative control; positive control; proximal inference

1. Introduction

Observational studies offer an important source of data for causal inference in socioeconomic, biomedical, and epidemiological research. A major challenge for observational studies is the potential for confounding factors of the exposure-outcome relationship in view. The impact of observed confounders on causal inference can be alleviated by direct adjustment methods such as inverse probability weighting, matching, regression, and doubly robust methods (Bang & Robins, 2005; Rosenbaum & Rubin, 1983b; Rubin, 1973; Stuart, 2010). However, unmeasured confounding is present in many observational studies. In this case, causal effects cannot be uniquely determined by the observed data without extra assumptions. As a result, the aforementioned adjustment methods may be biased and potentially misleading in the presence of unmeasured confounding. Sensitivity analysis methods (Cornfield et al., 1959; Rosenbaum & Rubin, 1983a) are widely used to evaluate the impact of unmeasured confounding and to assess the robustness of causal inferences, but in general they cannot completely correct for confounding bias. Auxiliary variables are particularly useful to adjust for unmeasured confounding in observational studies. The instrumental variable (IV) approach (Angrist et al., 1996; Baker & Lindeman, 1994; Goldberger, 1972; Robins, 1994; Wright, 1928), rests on an auxiliary covariate that (i) has no direct effect on the outcome, (ii) is independent of the unmeasured confounder, and (iii) is associated with the exposure. In addition, a structural outcome model or a monotone effect of the IV on the treatment, is typically required to identify a causal effect. Although the IV approach has gained popularity in causal inference literature in recent years, particularly in health and social sciences, the approach is highly sensitive to violation of any of assumptions (i)–(iii).

A more recent framework that leverages both negative control exposures and negative control outcomes to mitigate confounding bias is known as proximal (or negative control) inference (e.g., Miao, Geng, et al., 2018; Tchetgen Tchetgen et al., 2020), where a negative control outcome is an outcome variable that is associated with the confounder but not causally affected by the primary exposure, and a negative control exposure is an exposure variable that is correlated with the primary exposure or the confounder but does not causally affect the outcome of interest. Starting from an earlier version of this paper (Miao, Shi, et al., 2018), we develop an outcome confounding bridge framework for identification and inference about causal effects by using a pair of negative control exposure and outcome to account for unmeasured confounding bias. This line of work contributes to the literature by

CONTACT Wang Miao mwfy@pku.edu.cn Department of Probability and Statistics, Peking University, Beijing 100871, People's Republic of China

Supplemental data for this article can be accessed online at http://dx.doi.org/10.1080/24754269.2024.2390748.

providing an alternative identification assumption, proposing practical inference methods, and establishing connections to conventional approaches for confounding bias adjustment. A subsequence of related methods have been developed for categorical cases (Shi et al., 2020), dynamic treatment regime (Qi et al., 2022), heterogeneous treatment effect (Sverdrup & Cui, 2023), longitudinal studies (Ying et al., 2023), mediation analysis (Dukes et al., 2021), outcome-dependent sampling (Li et al., 2022), reinforcement learning (Bennett & Kallus, 2021), semiparametric theory (Cui et al., 2023), survival analysis (Ying et al., 2022), and synthetic control (Shi et al., 2021).

In this paper, we illustrate the outcome confounding bridge framework for identification and inference about causal effects. Our approach is based on a key assumption that the confounding effect on the primary outcome matches that on a transformation of the negative control outcome; throughout, this transformation is referred to as an outcome confounding bridge function, which is formally introduced in Section 3. Although in practice the bridge function is unknown, it can be identified by using a negative control exposure under certain completeness condition. Consistent and asymptotically normal estimation of the average causal effect can be achieved by the generalized method of moments described in Section 4. In Section 5, we generalize the negative control approach by allowing for a positive control outcome, which can be causally affected by the primary exposure. We also develop a sensitivity analysis approach for checking the robustness of causal inference against the negative control assumptions. In Section 6, we conduct simulation studies to evaluate the performance of the double negative control approach and compare it to competing methods. In Section 7, we apply our approach to a time-series study about the effect of air pollution on mortality. We conclude in Section 8 with a discussion about the implications of our approach in observational studies and modern data science.

2. Definition and examples of negative control outcomes

Throughout, we let X, Y, and V denote the primary exposure, outcome, and a vector of observed covariates, respectively. Vectors are assumed to be column vectors, unless explicitly transposed. Following the convention in causal inference, we use $Y(x)$ to denote the potential outcome under an intervention that sets X to x, and maintain the consistency assumption that the observed outcome is a realization of the potential outcome under the exposure actually received: $Y = Y(x)$ when $X = x$. We focus on the average causal effect (ACE) of X on Y, which is a contrast of the potential outcome mean between two exposure levels, for instance, $\text{ACE}_{XY} = E\{Y(1) - Y(0)\}$ for a binary exposure.

The ignorability assumption stating that $Y(x) \perp\!\!\!\perp X \mid V$ is conventionally made in causal inference, but it does not hold when unmeasured confounding is present. In this case, latent ignorability that states $Y(x) \perp\!\!\!\perp X \mid (U, V)$ is more reasonable, allowing for an unobserved confounder U. For notational convenience, we present results conditionally on observed covariates and suppress V unless otherwise stated.

Assumption 2.1 (Latent ignorability): $Y(x) \perp\!\!\!\perp X \mid U$ *for all* x.

Given latent ignorability, we have that for all x,

$$E\{Y(x)\} = E\{E(Y \mid U, X = x)\}. \tag{1}$$

The crucial difficulty of implementing (1) is that U is not observed and both the conditional mean $E(Y \mid U, X = x)$ and the density function $f(U)$ are not identified.

We introduce negative control variables to mitigate the problem of unmeasured confounding. Suppose an auxiliary outcome W is available and satisfies the following assumption.

Assumption 2.2 (Negative control outcome): $W \perp\!\!\!\perp X \mid U$ *and* $W \not\perp\!\!\!\perp U$.

The assumption realizes the notion of a negative control outcome that it is associated with the confounder but not causally affected by the primary exposure. Moreover, the sets of unmeasured confounders for (X, W) and (X, Y) are the same, which corresponds to the U-comparable assumption of Lipsitch et al. (2010). Assumption 2.2 does not impose restrictions on the association of W–Y. A special case is the nondifferential assumption of Lipsitch et al. (2010) and Tchetgen Tchetgen (2014), which further requires $W \perp\!\!\!\perp Y \mid U$ and does not allow for extra confounders of W–Y association. Justification of Assumption 2.2 and choice of negative controls require subject matter knowledge.

Example 2.1: In a study about the effect of acute stress on mortality from heart disease, Trichopoulos et al. (1983) found increasing mortality from cardiac and external causes during the days immediately after the 1981 earthquake

in Athens. However, acute stress due to the earthquake is unlikely to quickly cause deaths from cancer. In a parallel analysis, they found no increase in risk of cancer mortality, which is evidence in favour of no confounding and reinforces their claim that acute stress increases mortality from heart diseases. In this study, the exposure is the psychological stress post-earthquake, and the outcome of interest is deaths from cardiac events. The unmeasured confounder can be the nutritional level or economic status which influences both the exposure and outcome. However, the exposure does not directly influence other causes of death such as cancer. The cancer mortality is the negative control outcome, and they are used to test whether confounding bias is present and to evaluate the plausibility of a causal association.

However, it is far more challenging to identify a causal effect with a single negative control outcome. In the Supplementary Material, we provide two distinct parameter values of a fully parametric model that lead to the identical distribution of (X, Y, W). In the next section, we explore more realistic conditions under which identification can be achieved.

3. Identification of causal effects with a negative control pair

The literature on proximal causal inference indicates it is possible to identify the average treatment effect with both a negative control outcome and a negative control exposure, by invoking an outcome confounding bridge function (Cui et al., 2023; Miao, Geng, et al., 2018) to characterize relationship between the confounding effects on both the primary outcome and the negative control outcome.

Assumption 3.1 (Outcome confounding bridge): *There exists some function* $b(W, X)$ *such that for all* x*,*

$$E(Y \mid U, X = x) = E\{b(W, x) \mid U, X = x\}. \tag{2}$$

When covariates V are observed, (2) becomes $E\{Y \mid U, V, X = x\} = E\{b(W, V, x) \mid U, V, X = x\}$. Assumption 3.1 states that the confounding effect of U on Y at exposure level x, is equal to the confounding effect of U on the variable $b(W, x)$, a transformation of W; it goes beyond U–comparability by characterizing the relationship between the confounding effects of U on Y and W. We illustrate the assumption with an example of the linear outcome confounding bridge. Assuming that $E(Y \mid U, X) = (1, X, U, XU)\beta$ and that $E(W \mid U)$ is linear in U, then (2) holds with $b(W, X; \gamma) = (1, X, W, XW)\gamma$, for an appropriate value of γ. Linearity in W in this bridge function, corresponds to a proportional relationship between the confounding effects of U on Y and W.

The average causal effect can be recovered by integrating the outcome confounding bridge over W.

Proposition 3.1: *Given Assumptions* 2.1–3.1*, we have that for all* x*,*

$$E\{Y(x)\} = E\{b(W, x)\}. \tag{3}$$

The proposition reveals the role of the negative control outcome and the outcome confounding bridge $b(w, x)$. Given $b(w, x)$, the potential outcomes mean and the average causal effect can be identified without an additional assumption. We emphasize that without knowledge of such bridge function, identification is not possible in general, even under a fully parametric model and full knowledge of the confounder distribution. However, in practice, the outcome confounding bridge is unknown. In order to identify the confounding bridge, we introduce an auxiliary exposure variable named negative control exposure Z that satisfies the following exclusion restrictions.

Assumption 3.2 (Negative control exposure): $Z \perp\!\!\!\perp Y \mid (U, X)$ *and* $Z \perp\!\!\!\perp W \mid (U, X)$.

The assumption states that upon conditioning on the primary exposure and the confounder, Z does not affect either the primary outcome Y nor the negative control outcome W. This assumption does not impose restrictions on the association between Z and X and allows Z to be confounded. A special case is the instrumental variable (Goldberger, 1972; Wright, 1928) that is independent of the confounder, in addition to the exclusion restrictions. Below we provide an empirical example for negative control exposures.

Example 3.2: In a time-series study about air pollution, Flanders et al. (2017) used air pollution level in future days as negative control exposures to test and reduce confounding bias. For day i, let X_i, Y_i, U_i denote the air pollution level (e.g., PM2.5), a public health outcome (e.g., mortality), and the unmeasured confounder, respectively; although Y_i is possibly affected by air pollution in the current and past days, it is not affected by future days air pollution,

X_{i+1} for instance; moreover, public health outcomes in general do not affect air pollution in the immediate future. Thus, it is reasonable to use X_{i+1} as a negative control exposure.

Just as negative control outcomes, a negative control exposure can also be used to test whether confounding bias occurs by checking if Z is independent of Y or W after conditioning on X. Alternatively, we propose to use a negative control exposure to identify the outcome confounding bridge. Taking expectation of U with respect to $f(U \mid Z, X)$ on both sides of $E(Y \mid U, X) = E\{b(W, X) \mid U, X\}$, we obtain

$$E(Y \mid Z, X) = E\{b(W, X) \mid Z, X\}. \tag{4}$$

The equation suggests that the outcome confounding bridge also captures the relationship between the crude effects of Z on Y and W. This is because conditional on X, the crude effects of Z on (Y, W) are completely driven by the association with the confounder U. Equation (4) offers a feasible strategy to identify the outcome confounding bridge with a negative control exposure. Because $E(Y \mid Z, X)$ and $f(W \mid Z, X)$ can be obtained from the observed data, one can solve the equation for the bridge function. This type of integral equation is known as the Fredholm integral equation of the first kind. Consider the case where both W and Z are binary, and then (4) becomes two linear equations with two unknown parameters. The following condition concerning the completeness of $f(W \mid Z, X)$ guarantees the uniqueness of the solution.

Assumption 3.3 (Completeness of $f(W \mid Z, X)$): *For all x, $W \not\perp\!\!\!\perp Z \mid X = x$; and for any square-integrable function g, if $E\{g(W) \mid Z = z, X = x\} = 0$ for almost all z, then $g(W) = 0$ almost surely.*

Completeness is a commonly-made assumption in identification problems, such as instrumental variable identification discussed by Newey and Powell (2003), D'Haultfœuille (2011), Darolles et al. (2011) and Andrews (2017). These previous results about completeness can equally be applied here. For a binary confounder, completeness holds as long as $W \not\perp\!\!\!\perp Z \mid X = x$ for all x; completeness also holds for many widely-used distributions such as exponential families (Newey & Powell, 2003) and location-scale families (Hu & Shiu, 2018). However, if ACE is of primary interest, the uniqueness assumption is not a prerequisite for estimation and inference, as indicated in Zhang et al. (2023).

Theorem 3.3: *Under Assumptions* 2.1–3.3, *Equation* (4) *has a unique solution, and the potential outcome mean is identified by plugging such solution into Equation* (3).

So far, under the completeness condition, we have identified the potential outcome mean without imposing any model restriction on the outcome confounding bridge. If the bridge function belongs to a parametric or semiparametric model, the completeness condition can be weakened.

Theorem 3.4: *Under Assumptions* 2.1–3.2 *and given a model $b(W, X; \gamma)$ for the bridge function indexed by a finite or infinite dimensional parameter γ, if for all x, $E\{b(W, x; \gamma) - b(W, x; \gamma') \mid Z, X = x\} \neq 0$ with a positive probability for any $\gamma \neq \gamma'$, then γ is identified by solving $E\{Y - b(W, X; \gamma) \mid Z, X\} = 0$, and thus the potential outcome mean is identified.*

For instance, the linear model $b(W, X; \gamma) = (1, X, W, XW)\gamma$ is identified as long as $E(W \mid Z, X) \neq E(W \mid X)$ with a positive probability, i.e., W is not mean independent of Z after conditioning on X. Under the linear outcome confounding bridge, the relationship between the causal effect, the confounding bias, and crude effects has an explicit form, as shown in the following example.

Example 3.5: Consider binary exposures (X, Z) and the linear confounding bridge function, $b(W, X; \gamma) = \gamma_0 + \gamma_1 X + \gamma_2 W + \gamma_3 XW$, and let $\mathrm{RD}_{XY \mid Z} = E(Y \mid X = 1, Z) - E(Y \mid X = 0, Z)$ denote the risk difference of X on Y conditional on Z; then (γ_2, γ_3) are identified by

$$\gamma_2 = \frac{\mathrm{RD}_{ZY \mid X=0}}{\mathrm{RD}_{ZW \mid X=0}}, \quad \gamma_3 = \frac{\mathrm{RD}_{ZY \mid X=1}}{\mathrm{RD}_{ZW \mid X=1}} - \gamma_2.$$

The average causal effect of X on Y is identified by

$$\begin{aligned} \mathrm{ACE}_{XY} = {} & E(\mathrm{RD}_{XY \mid Z}) - (\gamma_2 + \gamma_3) E(\mathrm{RD}_{XW \mid Z}) \\ & + \gamma_3 \sum_{z=0}^{1} \{\mathrm{RD}_{XW \mid Z=z} \times f(Z = z, X = 1)\}. \end{aligned}$$

If the bridge function is additive, i.e., assuming that $\gamma_3 = 0$, then $\gamma_2 = E(\mathrm{RD}_{ZY\,|\,X})/E(\mathrm{RD}_{ZW\,|\,X})$ and

$$\mathrm{ACE}_{XY} = E(\mathrm{RD}_{XY\,|\,Z}) - \frac{E(\mathrm{RD}_{ZY\,|\,X})}{E(\mathrm{RD}_{ZW\,|\,X})} \times E(\mathrm{RD}_{XW\,|\,Z}). \tag{5}$$

This example offers a convenient adjustment when only summary data about crude effects are available. In the Supplementary Material, we extend this example by allowing for exposures of arbitrary type and a nonparametric outcome confounding bridge. Identification of causal effect is also possible without completeness condition, see (Zhang et al., 2023).

So far, we have identified the average causal effect with a pair of negative control exposure and outcome. If the treatment effect on the treated, $E\{Y(1) - Y(0) \mid X = 1\}$, is of interest instead, one only needs a weakened outcome confounding bridge assumption imposed on the control group, i.e., $E(Y \mid U, X = 0) = E\{b(W) \mid U, X = 0\}$ for some function $b(W)$, and then a negative control exposure can be used to identify $b(W)$. Our confounding bridge approach clarifies the roles of negative control exposure and outcome in confounding bias adjustment. A negative control outcome is used to mimic unobserved potential outcomes via the outcome confounding bridge that captures the relationship between the effects of confounding. The confounding bridge approach unifies previous bias adjustment methods in the negative control design. The approaches of Tchetgen Tchetgen (2014) and Sofer et al. (2016) are special cases of our outcome confounding bridge approach by assuming rank preservation of individual potential outcomes or monotonicity about the confounding effects. The factor analysis approach of Gagnon-Bartsch et al. (2013) and Wang et al. (2017) in fact identifies the outcome confounding bridge via factor loadings on the confounder. Therefore, these previous approaches reinforce the key role of the confounding bridge in the negative control design. Confounder proxies used by Miao, Geng, et al. (2018) and Kuroki and Pearl (2014) can be viewed as special negative controls in our framework. The identification strategy of Miao, Geng, et al. (2018) rests on a completeness condition involving the unmeasured confounder; however, our completeness condition depends only on observed variables. Our identification strategy rests on the outcome confounding bridge; alternatively, Cui et al. (2023) propose an identification approach that rests on an exposure confounding bridge $e(Z, X)$ defined by the solution to $E\{e(Z, X) \mid W, X = x\} = \{p(X = x \mid W)\}^{-1}$, connecting the negative control exposure to the inverse propensity score. Their identification is guaranteed by a completeness condition of $p(U \mid W, X)$.

4. Estimation

We focus on the estimation of the average causal effect $\Delta = E\{Y(x_1) - Y(x_0)\}$ that contrasts potential outcomes mean under two exposure levels x_1 and x_0. We first consider estimation with i.i.d. data samples and then generalize to time-series data. Suppose that one has specified a parametric model for the outcome confounding bridge, $b(W, V, X; \gamma)$. Practically, we recommend users start with a linear additive $b(W, X) = \gamma_1 W + \gamma_2 X$ or exponential multiplicative $b(W, X) = \exp(\gamma_1 W + \gamma_2 X)$. However, a misspecified low dimensional model $b(W, X)$ can potentially lead to a biased result. The users can use a variety of more flexible approaches such as semiparametric (e.g., partially linear model, single index model) or nonparametric (e.g., generalized additive, reproducing kernels, neural networks, see Cui et al., 2023; Kallus et al., 2021), to check the robustness of the estimated causal effect on $b(W, X)$, thus further alleviating concerns about misspecification bias. A standard approach to estimate $\theta = (\gamma, \Delta)$ is the generalized method of moments (Hall, 2005; Hansen, 1982). We let $D_i = (X_i, Z_i, Y_i, W_i, V_i), 1 \le i \le n$ denote the observed data samples. Define the vector of moment restrictions

$$h(D_i; \theta) = \begin{cases} \{Y_i - b(W_i, V_i, X_i; \gamma)\} \times q(X_i, V_i, Z_i), \\ \Delta - \{b(W_i, V_i, x_1; \gamma) - b(W_i, V_i, x_0; \gamma)\}, \end{cases} \tag{6}$$

with a user-specified vector function q, and let $m_n(\theta) = 1/n \sum_{i=1}^{n} h(D_i; \theta)$; the GMM solves $\widehat{\theta} = \arg\min_\theta m_n^\top(\theta)$ $\Omega\, m_n(\theta)$, with a user-specified positive-definite weight matrix Ω.

Typically, the dimension of q must be at least as large as that of γ. For instance, if $b(W, V, X; \gamma) = (1, X, V^\top, W)\gamma$, one can use $q(X, V, Z) = (1, X, V^\top, Z)^\top$ for the GMM. Cui et al. (2023) develop the semiparametric theory for double negative controls by assuming the existence of both the negative control outcome bridge function $b(W, X)$ and negative control exposure bridge function $e(Z, X)$. Their semiparametric efficient estimator is partially based on the above negative control estimator (6) from an unpublished initial draft of the current paper.

The GMM can equally be applied to time-series data for parameter estimation (Hamilton, 1994, chapter 14). Consider a typical time-series model,

$$Y_i = \gamma_0 + \gamma_1 X_i + U_i + \varepsilon_{1i}, \quad X_i = \alpha_0 + \alpha_1 U_i + \varepsilon_{2i}, \quad U_i = \xi U_{i-1} + (1 - \xi^2)^{1/2} \varepsilon_{3i},$$

with normal white noise $\varepsilon_{1i}, \varepsilon_{2i}, \varepsilon_{3i}$. As suggested by Flanders et al. (2017), $Z_i = X_{i+1}$ can be used as a negative control exposure; in addition, we use $W_i = Y_{i-1}$ as a negative control outcome, which satisfies $Z_i \perp\!\!\!\perp (W_i, Y_i) \mid$

(X_i, U_i) and $W_i \perp\!\!\!\perp X_i \mid U_i$. To estimate γ_1 via the GMM, we specify a linear outcome confounding bridge model $b(W_i, X_i, X_{i-1}; \gamma) = (1, X_i, X_{i-1}, W_i)\gamma$ and use $q(X_i, X_{i-1}, Z_i) = (1, X_i, X_{i-1}, Z_i)^\top$ to construct the moment restrictions. It seems surprising that we can consistently estimate γ_1 when we only observe X and Y but not U. However, this is achieved by selecting appropriate negative control exposure and outcome variables from the observed data for each observation. This approach benefits from the serial correlation of the confounder, but does not apply to independent observations.

Consistency and asymptotic normality of the GMM estimator have been established under appropriate conditions in Hansen (1982) and Hall (2005). Standard errors and confidence intervals can be constructed based on the normal approximation,

$$n^{1/2}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma_1 \Sigma_0 \Sigma_1^\top),$$

where θ_0 denotes the true value of θ, and

$$\Sigma_1 = (M^\top \Omega M)^{-1} M^\top \Omega, \quad M = \lim_{n \to +\infty} \frac{\partial m_n(\theta)}{\partial \theta^\top}\bigg|_{\theta=\theta_0}, \quad \Sigma_0 = \lim_{n \to +\infty} \mathrm{Var}\{n^{1/2} m_n(\theta_0)\}.$$

For i.i.d. data, a consistent estimator of the asymptotic variance can be constructed by using

$$\begin{aligned} \widehat{\Sigma}_1 &= (\widehat{M}^\top \Omega \widehat{M})^{-1} \widehat{M}^\top \Omega, \quad \widehat{M} = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial h(D_i; \theta)}{\partial \theta^\top}\bigg|_{\theta=\widehat{\theta}}, \\ \widehat{\Sigma}_0 &= \frac{1}{n} \sum_{i=1}^{n} h(D_i; \widehat{\theta}) h^\top(D_i; \widehat{\theta}); \end{aligned} \tag{7}$$

and a 95% confidence interval for the elements of θ in large samples is $\widehat{\theta} \pm 1.96 \times \{\mathrm{diag}(\widehat{\Sigma}_1 \widehat{\Sigma}_0 \widehat{\Sigma}_1^\top)/n\}^{1/2}$, where diag denotes the diagonal elements of a matrix. Variance estimation in the time-series setting is more complicated due to the serial correlation. When the observed data are serially correlated, $\widehat{\Sigma}_0$ in (7) is no longer consistent for Σ_0, and one should use heteroscedasticity and autocorrelation covariance (HAC) estimators that are consistent under relatively weak assumptions (Andrews, 1991; Newey & West, 1987). In this paper, we use the Newey-West estimate of Σ_0:

$$\begin{aligned} \Sigma_0^{\mathrm{HAC}} &= \widehat{\Sigma}_0 + \sum_{i=1}^{b_n} \left\{1 - \frac{i}{1 + b_n}\right\} (\widehat{\Sigma}_i + \widehat{\Sigma}_i^\top), \quad b_n = c \times n^{1/3} \text{ for some constant c}, \\ \widehat{\Sigma}_i &= \frac{1}{n} \sum_{j=i+1}^{n} h(D_j; \widehat{\theta}) h^\top(D_{j-i}; \widehat{\theta}), \end{aligned}$$

where b_n is the bandwidth parameter controlling the number of auto-covariances included in the HAC estimator; for practical guidance for the choice of b_n, see (Andrews, 1991) and (Hall, 2005, Section 3.5.3). In contrast to the i.i.d. setting, the HAC estimator includes extra covariance terms $\{\widehat{\Sigma}_i, i \neq 0\}$ to account for the serial correlation.

5. Positive control outcome

The negative control outcome assumption, $W \perp\!\!\!\perp X \mid U$, is not met when the auxiliary outcome W is causally affected by X. In this case, we call W a positive control outcome. Let $W(x)$ denote the potential outcome of W when X is set to x; the following assumption preserves U-comparability but accommodates a nonzero causal effect of X on W, see Figure 1 for a DAG model.

Assumption 5.1 (Positive control outcome): $W(x) \perp\!\!\!\perp X \mid U$ *for all* x.

Proposition 5.1: *Given the latent ignorability Assumption* 2.1*, the outcome confounding bridge Assumption* 3.1*, and the positive control Assumption* 5.1*, then* $E\{Y(x)\} = E\{b(W(x), x)\}$ *for all* x.

The potential outcome mean $E\{Y(x)\}$ depends on the distribution of $W(x)$ rather than the observed distribution of W. Given a positive control outcome and a negative control exposure, (4) still holds, and thus can be used to identify the outcome confounding bridge. As a consequence, the causal effect of X on Y can be identified if both a positive control outcome and a negative control exposure are available and the causal effect of X on

W is known a priori. Suppose the bridge function has an additive form $b(W(X), X; \gamma) = b_1(X; \gamma_1) + b_2(W(X))$ where the structural parameter γ_1 is unknown. Then, the potential outcome mean can be rewritten as $E\{Y(x)\} = E\{b_1(X; \gamma_1) + b_2(W(X))\}$. We let $\gamma_2(x) = E\{b_2(W(x))\}$ be a specified functional form of x in a sensitivity analysis measuring the mean of $W(x)$ transformed by some function b_2. The estimation is analogous to the GMM method in Section 4. Define the vector of moment restrictions

$$h(X_i, Y_i; \gamma_1, \gamma_2(x), \Delta) = \begin{cases} [Y_i - b_1(X_i; \gamma_1) - \gamma_2(X_i)] \times q(X_i), \\ \Delta - [b_1(x_1; \gamma_1) - b_1(x_0; \gamma_1) + \gamma_2(x_1) - \gamma_2(x_0)], \end{cases} \tag{8}$$

with a user-specified vector function q. The first component in (8) consists of unbiased estimating equations for γ_1 because $E\{Y - b_1(X) - \gamma_2(X) \mid X\} = 0$, and the second one for Δ. In practice, the users can make use of auxiliary information of $\gamma_2(x)$ if possible or specify a functional form based on expert knowledge to test the robustness of the estimation method against the effect size on the positive control. We further illustrate this with the following examples.

Example 5.2: Consider binary exposures (X, Z) and the linear outcome confounding bridge $b(W(X), X) = \gamma_0 + \gamma_1 X + \gamma_2 W(X)$ for a positive control outcome W. Then $E\{Y(x)\} = \gamma_0 + \gamma_1 x + \gamma_2 E\{W(x)\}$ and $\text{ACE}_{XY} = \gamma_1 + \gamma_2 \times \text{ACE}_{XW}$. Identification of (γ_1, γ_2) is identical as in the negative control outcome case, with $\gamma_2 = E(\text{RD}_{ZY\,|\,X})/E(\text{RD}_{ZW\,|\,X})$ and $\gamma_1 = E(\text{RD}_{XY\,|\,Z}) - \gamma_2 \times E(\text{RD}_{XW\,|\,Z})$. In contrast with the negative control setting in Example 3.5, identification with a positive control outcome involves the average causal effect of X on W. Using ACE_{XW} as a sensitivity parameter, sensitivity analysis can be performed to evaluate the plausibility of a causal effect of X on Y; if ACE_{XW} is known to belong to the interval $[a, b]$, then the bound for ACE_{XY} is $[\gamma_1 + \gamma_2 a, \gamma_1 + \gamma_2 b]$; given the sign of γ_2, the sign of $E(\text{RD}_{XY\,|\,Z}) - \text{ACE}_{XY}$, i.e., the confounding bias, can be inferred from the sign of $E(\text{RD}_{XW\,|\,Z}) - \text{ACE}_{XW}$.

Example 5.3: In studies assessing the effect of intrauterine smoking (X) on offspring birthweight (Y) and seven years old body mass index (W), Davey Smith (2008) and Davey Smith (2012) used paternal smoking (Z) as a negative control exposure, and observed that

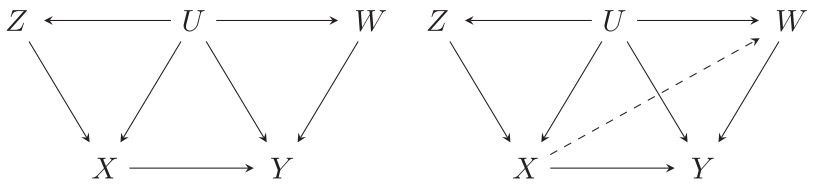
$$E(\text{RD}_{XY\,|\,Z}) = -150\,\text{g}, \quad E(\text{RD}_{XW\,|\,Z}) = 0.15\,\text{kg/m}^2,$$
$$E(\text{RD}_{ZY\,|\,X}) = -10\,\text{g}, \quad E(\text{RD}_{ZW\,|\,X}) = 0.11\,\text{kg/m}^2.$$

Following the analysis in Example 5.2, we obtain $\gamma_2 = -91, \gamma_1 = -136$, and thus $\text{ACE}_{XY} = -136 - 91 \times \text{ACE}_{XW}$g. A necessary condition to explain away the observed impact of intrauterine smoking on birthweight (i.e., to make $\text{ACE}_{XY} \geq 0$) is $\text{ACE}_{XW} \leq -1.5\ \text{kg/m}^2$, a protective effect of intrauterine smoking on later-life body mass index. However, intrauterine smoking is unlikely to have such a considerable protective effect against obesity, and in fact, researchers have hypothesized although not definitely established that intrauterine smoking is likely to increase not decrease the risk of offspring obesity (Mamun et al., 2006). Therefore, the most plausible explanation is that intrauterine smoking decreases offspring birthweight, at least −136 g on average if one believes intrauterine smoking can also cause offspring adiposity.

6. Simulation studies

6.1. Simulations for a binary exposure

We provide two simulation examples in this and the next section. In the first simulation, we generate two variables $V, U \sim N(0, 1)$ with correlation $\sigma_{uv} = 0.5$. Then we generate the negative control exposure, negative control outcome based on the following models $Z = 0.5 + 0.5V + U + \varepsilon_1$, $W = 1 - V + \xi U + \varepsilon_2$ with $\varepsilon_1, \varepsilon_2 \sim N(0, 1)$.



(a) Negative control (b) Positive control

Figure 1. DAG models for negative and positive controls. The dashed arrow indicates a possibly nonzero causal effect of X on W. (a) Negative control and (b) Positive control.

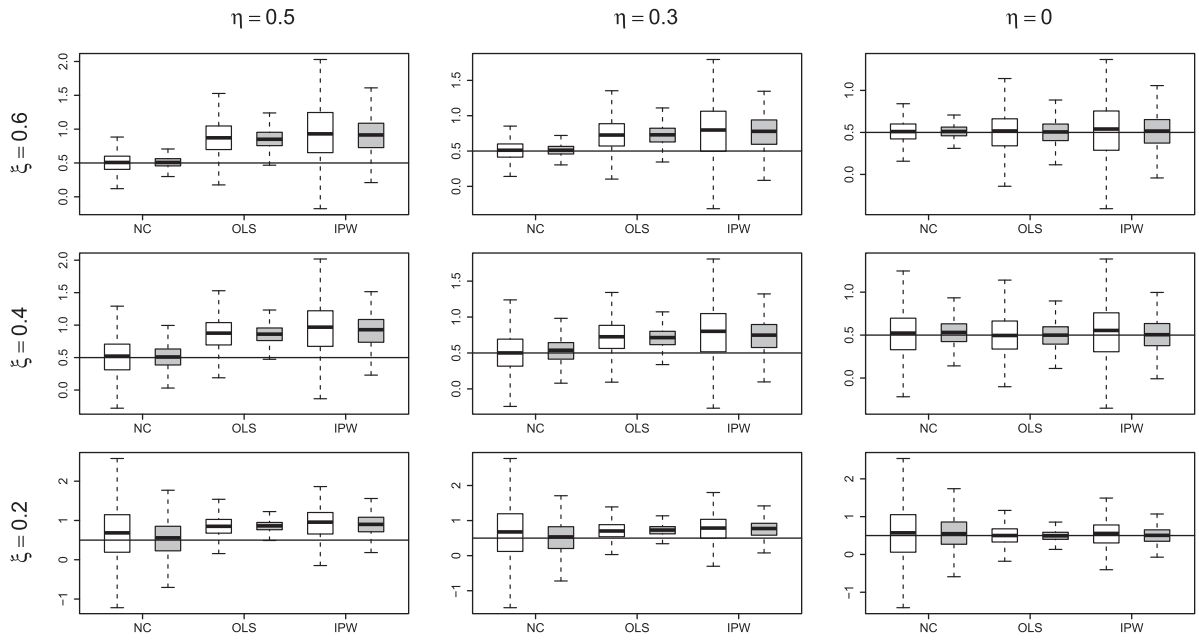


Figure 2. Boxplots for estimators of the average causal effect.
Note: For NC, $b = (1, X, V, W, XV, XW)\gamma$ and $q = (1, X, V, Z, XV, XZ)^\top$ are used for the GMM; for IPW, a logistic model for $f(X = 1 \mid V)$ is used; for OLS, a linear outcome model is used. White boxes are for a sample size 500 and grey ones 1500; the horizontal line marks the true value of the average causal effect.

Table 1. Coverage probability of 95% negative control confidence interval for the average causal effect.

		$\eta = 0.5$		0.3		0	
	$n =$	500	1500	500	1500	500	1500
$\xi =$	0.6	0.945	0.936	0.958	0.953	0.954	0.935
	0.4	0.958	0.957	0.968	0.955	0.964	0.956
	0.2	0.953	0.963	0.970	0.963	0.978	0.979

The exposure and the potential outcome are generated based on $\text{logit}\{f(X = 1 \mid Z, V, U)\} = -0.5 + Z + 0.5V + \eta U$, $Y(x) = 1 + 0.5x + 2V + U + 1.5xU + 2\varepsilon_2$ with η encoding the magnitude of confounding and ξ the association between the negative control outcome and the confounder. We analyse data with the negative control approach (NC), standard inverse probability weighting (IPW), and ordinary least square (OLS).

For each choice of $\eta = 0, 0.3, 0.5$ and $\xi = 0.2, 0.4, 0.6$, we replicate 1000 simulations at sample size 500 and 1500, respectively, and summarize results as boxplots in Figure 2. From Figure 2, the negative control estimator has a small bias in all settings; in contrast, ordinary least square and inverse probability weighted estimators are biased except under no unmeasured confounding ($\eta = 0$). When the association between the negative control outcome and the confounder is moderate to strong ($\xi = 0.4, 0.6$), the negative control estimator is more efficient than the other two, but has greater variability otherwise ($\xi = 0.2$). Table 1 presents coverage probabilities of 95% negative control confidence intervals based on a normal approximation, which generally approximate the nominal level of 0.95. But, when the association between the negative control outcome and the confounder is weak ($\xi = 0.2$), the coverage probabilities are slightly inflated. Therefore, we recommend the negative control approach to remove the confounding bias in observational studies, and to enhance efficiency, we recommend when possible using a negative control outcome that is strongly associated with the confounder.

6.2. Simulations for time series data

We generate time-dependent data according to

$$U_i = \xi U_{i-1} + (1 - \xi^2)^{1/2}\varepsilon_{1i}, \quad V_i = 0.6U_i + \varepsilon_{2i}, \quad X_i = 0.4 + 1.5V_i + \eta U_i + \varepsilon_{3i},$$
$$Y_i = 0.5 + 0.7X_i + 1.5V_i + 0.9U_i + \varepsilon_{4i}, \quad \varepsilon_{1i}, \varepsilon_{2i}, \varepsilon_{3i}, \varepsilon_{4i} \sim N(0, 1),$$

where U_i is a stationary autoregressive process with autocorrelation coefficient ξ, and η controls the magnitude of confounding. We analyse data with the negative control approach (NC), ordinary least square (OLS) without

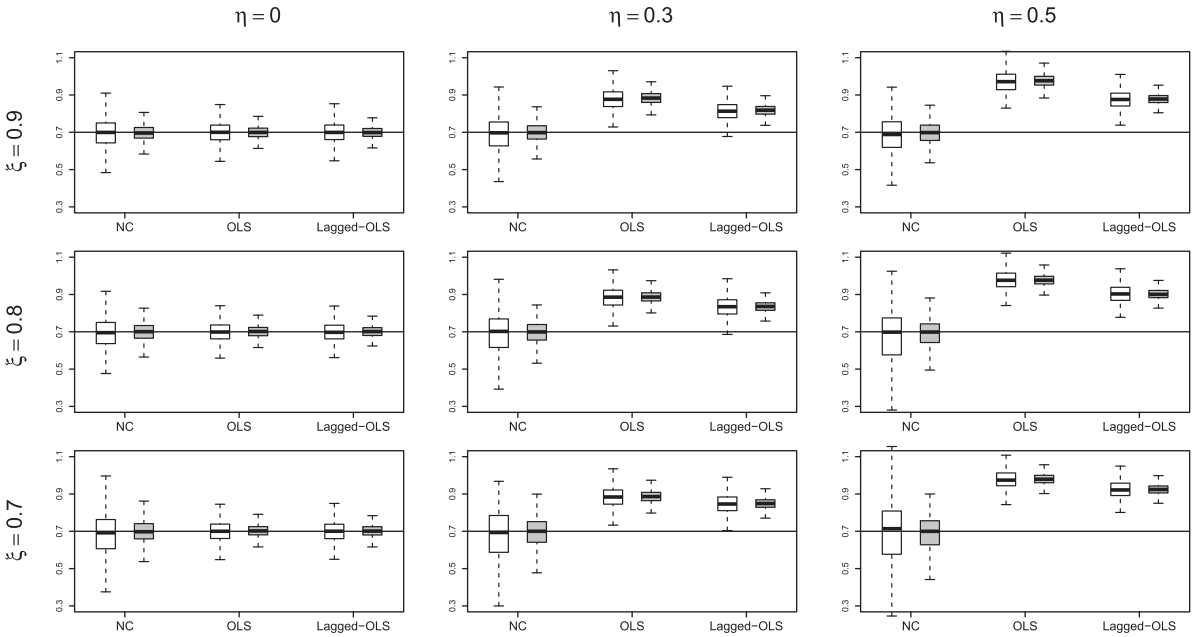


Figure 3. Boxplots for time series data analysis.
Note: For NC, $b = (1, X_i, X_{i-1}, V_i, V_{i-1}, W_i)\gamma$ and $q = (1, X_i, X_{i-1}, V_i, V_{i-1}, Z_i)^\top$ are used for the GMM. White boxes are for a sample size 500 and grey ones 1500; the horizontal line marks the true value of the structural parameter.

Table 2. Coverage probability of 95% negative control confidence interval for the time-series model.

		$\eta = 0$		0.3		0.5	
	$n =$	500	1500	500	1500	500	1500
$\xi =$	0.9	0.953	0.947	0.948	0.950	0.950	0.947
	0.8	0.979	0.952	0.952	0.943	0.933	0.946
	0.7	0.982	0.974	0.937	0.942	0.912	0.940

controlling lagged exposures, and lagged-OLS by controlling one-day lagged exposure. For the negative control approach, we use $W_i = Y_{i-1}$ and $Z_i = X_{i+1}$ as negative controls, and do not need auxiliary data.

For each choice of $\xi = 0.7, 0.8, 0.9$ and $\eta = 0, 0.3, 0.5$, we replicate 1000 simulations at sample size 500 and 1500, respectively. Figure 3 presents boxplots of the estimators. The negative control estimator has a small bias in all nine scenarios, and its variability becomes smaller as the autocorrelation of the confounder process increases. The 95% negative control confidence intervals based on the Newey and West (1987) variance estimator have coverage probability approximating 0.95, as shown in Table 2. The ordinary least square estimator is biased except under no unmeasured confounding ($\eta = 0$), in which case, it is more efficient than the negative control estimator. Controlling lagged exposures in ordinary least squares can reduce confounding bias, but cannot eliminate it. Therefore, we recommend the negative control approach for the estimation of a linear time-series regression model in the presence of unmeasured confounding.

7. Evaluation of the effect of air pollution on mortality

While there are many long-term threats posed by air pollution, its acute effects on mortality also pose an important public health concern. We apply the negative control approach to evaluate the short-term effect of air pollution on mortality using datasets from a time-series study in Philadelphia, New York, and Boston. Here we present the analysis results for Philadelphia and relegate those for the other two cities to the Supplementary Material. The dataset for Philadelphia contains $n = 2621$ daily records of PM2.5, temperature, ozone, date, and number of deaths in Philadelphia from 1999 to 2006. With accidental deaths excluded, the number of deaths ranges from 73 to 179, which is often assumed to have a Poisson distribution. In our analysis, we use the square root of the number of deaths for the purpose of normalization and variance stabilization (Freeman & Tukey, 1950).

For a given day i, we let Y_i denote the square root of number of deaths, X_i be the PM2.5 concentration measurement, V_i consist of temperature and its square, ozone, and X_{i-1} to control lagged effects, and T_i consist of polynomial and Fourier bases of time to account for both secular and seasonal trends:

$$T_i = \{i/n, i^2/n^2, \sin(2\pi i/365), \cos(2\pi i/365), \ldots, \sin(8\pi i/365), \cos(8\pi i/365)\}.$$

Table 3. Estimates and 95% confidence intervals (×10,000) of the effect of air pollution in Philadelphia.

	Number of lagged exposures controlled					
	One day		Two days		Three days	
	Estimate	*p*-value	Estimate	*p*-value	Estimate	*p*-value
			Ordinary least square			
β_1	84 (48, 120)	0	78 (41, 115)	0	79 (43, 116)	0
			Confounding test			
α_1	−40 (−73, −7)	0.0167	−39 (−71, −7)	0.0174	−40 (−72, −7)	0.0158
α_2	41 (11, 71)	0.0072	40 (10, 69)	0.0080	39 (10, 69)	0.0083
			Negative control estimation			
β_1	45 (−6, 97)	0.0854	46 (−6, 98)	0.0844	46 (−7, 99)	0.0915

We assume a linear outcome model, $Y_i = \beta_1 X_i + (1, V_i, T_i)\beta_2 + U_i$, and we are interested in the regression coefficient β_1 that encodes the immediate effect of current day PM2.5 on mortality. All results are summarized in Table 3, where confidence intervals and p-values are obtained from the normal approximation and the Newey and West (1987) variance estimator is used to account for serial correlation. A standard regression analysis shows that short-term exposure to PM2.5 can significantly increase mortality, with point estimate 0.0084 and 95% confidence interval $(0.0048, 0.0120)$ for β_1. However, a confounding test by fitting the model

$$W_i = \alpha_1 X_i + \alpha_2 Z_i + (1, X_{i-1}, V_{i-1}, T_{i-1})\alpha_3 + U_{i-1},$$

with $W_i = Y_{i-1}$, results in point estimate -0.0040 of α_1 with 95% confidence interval $(-0.0073, -0.0007)$ and p-value 0.0167, and point estimate 0.0041 of α_2 with 95% confidence interval $(0.0011, 0.0071)$ and p-value 0.0072. These results suggest the presence of unmeasured confounding because W_i occurs before X_i and Z_i, and should not be affected by them. Thus, the ordinary least squares method appears not entirely appropriate in this setting. We apply the proposed negative control approach and use $Z_i = X_{i+1}$ and $W_i = Y_{i-1}$ as the negative control exposure and outcome, respectively. We assume a linear outcome confounding bridge $b = (1, X_i, V_i, V_{i-1}, T_i, W_i)\beta$, and use $q = (1, X_i, V_i, V_{i-1}, T_i, Z_i)^\top$ for the GMM. Compared to the standard regression, the negative control estimate of β_1 is attenuated toward zero a lot, although it still has some significance with point estimate 0.0045 and 95% confidence interval $(-0.0006, 0.0097)$. Further analyses controlling longer lagged exposures by including X_{i-2} and X_{i-3} in V_i lead to analogous results as those obtained when only X_{i-1} is controlled. Our analyses indicate the presence of unmeasured confounding in the air pollution study in Philadelphia. In parallel analyses we provide in the Supplemental Materials, unmeasured confounding is also detected in the dataset for New York via the negative control approach, but not detected in the dataset for Boston. After accounting for unmeasured confounding, our negative control inference shows a significant acute effect of PM2.5 on mortality in Philadelphia, but such an effect is not detected in New York or Boston.

8. Discussion

We propose an outcome confounding bridge approach for negative control/proximal inference on causal effects. We clarify the key assumptions and the roles of negative control outcome and exposure, and discuss robustness and sensitivity of the approach. In the supplementary material, we provide some insights on the connection between the negative control and the instrumental variable approaches, focussing on the estimation of a structural model. As we illustrate, an invalid instrumental variable that fails to be independent of the unmeasured confounder can be viewed as a negative control exposure, and a negative control outcome can be used to repair such an invalid IV by applying our double negative control adjustment. Under a linear structural model, we show the double robustness property of the negative control estimator, in the sense that it is consistent if either the confounding bridge is correctly specified or the negative control exposure is a valid IV.

Besides causal effect evaluation, our approach has important implications for the design of observational studies. Even if an exposure or response factor is not directly relevant to the study variables in view, it is useful to collect them and use them as negative controls for the purpose of confounding diagnostic and bias adjustment. Time-series studies, such as the air pollution example we consider, are particularly well-suited for the proposed negative control approach, because negative controls can be constructed from observations of the exposure and outcome themselves. However, in general, our approach requires one to collect extra data about negative control variables.

The negative control assumptions we present in this paper describe the general principles for selecting negative control variables, and the examples we give provide guidance for certain specific studies; but in general, subject matter knowledge about the data-generating mechanism and the potentially unmeasured confounders, such as the specificity of the exposure-outcome relation (Hill, 1965; Lipsitch et al., 2010), is indispensable to choose an appropriate negative control.

Our approach has promising application in modern big and multi-source data analyses. Identification of the outcome confounding bridge and the average causal effect depends only on $f(Y, Z, X)$ and $f(W, Z, X)$ but not the joint distribution of (Y, W), and thus enjoys the convenience of data integration and two-sample inference. For certain outcome confounding bridge models such as the linear one, estimation of the average causal effect requires only summary but not individual-level data, and thus allows for synthetic analysis by using results from multiple studies. Such extensions will be carefully developed in the future.

Acknowledgments

We are grateful for valuable comments from the editor and two anonymous reviewers.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was partially supported by National Key R&D Program of China [grant numbers 2020YFE0204200, 2022YFA1008100] and National Natural Science Foundation of China [grant numbers 12071015, 12292983].

References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, *59*(3), 817–858. https://doi.org/10.2307/2938229

Andrews, D. W. K. (2017). Examples of L^2-complete and boundedly-complete distributions. *Journal of Econometrics*, *199*(2), 213–220. https://doi.org/10.1016/j.jeconom.2017.05.011

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*(434), 444–455. https://doi.org/10.1080/01621459.1996.10476902

Baker, S. G., & Lindeman, K. S. (1994). The paired availability design: A proposal for evaluating epidural analgesia during labor. *Statistics in Medicine*, *13*(21), 2269–2278. https://doi.org/10.1002/sim.v13:21

Bang, H., & Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics*, *61*(4), 962–973. https://doi.org/10.1111/biom.2005.61.issue-4

Bennett, A., & Kallus, N. (2021). Proximal reinforcement learning: Efficient off-policy evaluation in partially observed Markov decision processes. *Operations Research*, *72*(3), 1071–1086. https://doi.org/10.1287/opre.2021.0781

Cornfield, J., Haenszel, W., Hammond, E. C., Lilienfeld, A. M., Shimkin, M. B., & Wynder, E. L. (1959). Smoking and lung cancer: Recent evidence and a discussion of some questions. *Journal of the National Cancer Institute*, *22*(1), 173–203.

Cui, Y., Pu, H., Shi, X., Miao, W., & Tchetgen Tchetgen, E. (2023). Semiparametric proximal causal inference. *Journal of the American Statistical Association*, *119*(546), 1348–1359. https://doi.org/10.1080/01621459.2023.2191817

Darolles, S., Fan, Y., Florens, J. P., & Renault, E. (2011). Nonparametric instrumental regression. *Econometrica*, *79*(5), 1541–1565. https://doi.org/10.3982/ECTA6539

Davey Smith, G. (2008). Assessing intrauterine influences on offspring health outcomes: Can epidemiological studies yield robust findings? *Basic & Clinical Pharmacology & Toxicology*, *102*(2), 245–256. https://doi.org/10.1111/pto.2008.102.issue-2

Davey Smith, G. (2012). Negative control exposures in epidemiologic studies. *Epidemiology*, *23*(2), 350–351. https://doi.org/10.1097/EDE.0b013e318245912c

D'Haultfœuille, X. (2011). On the completeness condition in nonparametric instrumental problems. *Econometric Theory*, *27*(3), 460–471. https://doi.org/10.1017/S0266466610000368

Dukes, O., Shpitser, I., & Tchetgen Tchetgen, E. (2021). *Proximal mediation analysis*. arXiv preprint arXiv:2109.11904.

Flanders, W. D., Strickland, M. J., & Klein, M. (2017). A new method for partial correction of residual confounding in time-series and other observational studies. *American Journal of Epidemiology*, *185*(10), 941–949. https://doi.org/10.1093/aje/kwx013

Freeman, M. F., & Tukey, J. W. (1950). Transformations related to the angular and the square root. *The Annals of Mathematical Statistics*, *21*(4), 607–611. https://doi.org/10.1214/aoms/1177729756

Gagnon-Bartsch, J., Jacob, L., & Speed, T. P. (2013). *Removing unwanted variation from high dimensional data with negative controls* (Technical Report 820). Dept. Statistics, Univ. California.

Goldberger, A. S. (1972). Structural equation methods in the social sciences. *Econometrica*, *40*(6), 979–1001. https://doi.org/10.2307/1913851

Hall, A. R. (2005). *Generalized method of moments*. Oxford University Press.

Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, *50*(4), 1029–1054. https://doi.org/10.2307/1912775

Hill, A. B. (1965). The environment and disease: Association or causation?. *Proceedings of the Royal Society of Medicine*, *58*(5), 295–300. https://doi.org/10.1177/003591576505800503

Hu, Y., & Shiu, J.-L. (2018). Nonparametric identification using instrumental variables: Sufficient conditions for completeness. *Econometric Theory*, *34*(3), 659–693. https://doi.org/10.1017/S0266466617000251

Kallus, N., Mao, X., & Uehara, M. (2021). *Causal inference under unmeasured confounding with negative controls: A minimax learning approach*. arXiv:2103.14029.

Kuroki, M., & Pearl, J. (2014). Measurement bias and effect restoration in causal inference. *Biometrika*, *101*(2), 423–437. https://doi.org/10.1093/biomet/ast066

Li, K. Q., Shi, X., Miao, W., & Tchetgen Tchetgen, E. (2022). *Doubly robust proximal causal inference under confounded outcome-dependent sampling*. arXiv:2208.01237.

Lipsitch, M., Tchetgen Tchetgen, E., & Cohen, T. (2010). Negative controls: A tool for detecting confounding and bias in observational studies. *Epidemiology*, *21*(3), 383–388. https://doi.org/10.1097/EDE.0b013e3181d61eeb

Mamun, A. A., Lawlor, D. A., Alati, R., O'callaghan, M. J., Williams, G. M., & Najman, J. M. (2006). Does maternal smoking during pregnancy have a direct effect on future offspring obesity? Evidence from a prospective birth cohort study. *American Journal of Epidemiology*, *164*(4), 317–325. https://doi.org/10.1093/aje/kwj209

Miao, W., Geng, Z., & Tchetgen Tchetgen, E. (2018). Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, *105*(4), 987–993. https://doi.org/10.1093/biomet/asy038

Miao, W., Shi, X., & Tchetgen Tchetgen, E. (2018). *A confounding bridge approach for double negative control inference on causal effects*. arXiv: 1808.04945.

Newey, W. K., & Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, *71*(5), 1565–1578. https://doi.org/10.1111/ecta.2003.71.issue-5

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703–708. https://doi.org/10.2307/1913610

Qi, Z., Miao, R., & Zhang, X. (2022). Proximal learning for individualized treatment regimes under unmeasured confounding. *Journal of the American Statistical Association*, *119*(546), 915–928. https://doi.org/10.1080/01621459.2022.2147841

Robins, J. M. (1994). Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics-Theory and Methods*, *23*(8), 2379–2412. https://doi.org/10.1080/03610929408831393

Rosenbaum, P. R., & Rubin, D. B. (1983a). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society: Series B*, *45*(2), 212–218. https://doi.org/10.1111/j.2517-6161.1983.tb01242.x

Rosenbaum, P. R., & Rubin, D. B. (1983b). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55. https://doi.org/10.1093/biomet/70.1.41

Rubin, D. B. (1973). The use of matched sampling and regression adjustment to remove bias in observational studies. *Biometrics*, *29*(1), 185–203. https://doi.org/10.2307/2529685

Shi, X., Miao, W., Hu, M., & Tchetgen Tchetgen, E. (2021). *Theory for identification and inference with synthetic controls: A proximal causal inference framework*. arXiv:2108.13935.

Shi, X., Miao, W., Nelson, J. C., & Tchetgen Tchetgen, E. (2020). Multiply robust causal inference with double negative control adjustment for categorical unmeasured confounding. *Journal of the Royal Statistical Society: Series B*, *82*(2), 521–540. https://doi.org/10.1111/rssb.12361

Sofer, T., Richardson, D. B., Colicino, E., Schwartz, J., & E. J. Tchetgen Tchetgen (2016). On negative outcome control of unobserved confounding as a generalization of difference-in-differences. *Statistical Science*, *31*(3), 348–361. https://doi.org/10.1214/16-STS558

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, *25*(1), 1–21. https://doi.org/10.1214/09-STS313

Sverdrup, E., & Cui, Y. (2023). *Proximal causal learning of heterogeneous treatment effects*. arXiv:2301.10913.

Tchetgen Tchetgen, E. (2014). The control outcome calibration approach for causal inference with unobserved confounding. *American Journal of Epidemiology*, *179*(5), 633–640. https://doi.org/10.1093/aje/kwt303

Tchetgen Tchetgen, E. J., Ying, A., Cui, Y., Shi, X., & Miao, W. (2020). *An introduction to proximal causal learning*. arXiv:2009.10982.

Trichopoulos, D., Zavitsanos, X., Katsouyanni, K., Tzonou, A., & Dalla-Vorgia, P. (1983). Psychological stress and fatal heart attack: The athens (1981) earthquake natural experiment. *The Lancet*, *321*(8322), 441–444. https://doi.org/10.1016/S0140-6736(83)91439-3

Wang, J., Zhao, Q., Hastie, T., & Owen, A. B. (2017). Confounder adjustment in multiple hypothesis testing. *The Annals of Statistics*, *45*(5), 1863–1894. https://doi.org/10.1214/16-AOS1511

Wright, P. G. (1928). *Tariff on animal and vegetable oils*. Macmillan.

Ying, A., Cui, Y., & Tchetgen Tchetgen, E. (2022). *Proximal causal inference for marginal counterfactual survival curves*. arXiv:2204.13144.

Ying, A., Miao, W., Shi, X., & Tchetgen Tchetgen, E. J. (2023). Proximal causal inference for complex longitudinal studies. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *85*(3), 684–704. https://doi.org/10.1093/jrsssb/qkad020

Zhang, J., Li, W., Miao, W., & Tchetgen Tchetgen, E. (2023). Proximal causal inference without uniqueness assumptions. *Statistics & Probability Letters*, *198*, Article 109836. https://doi.org/10.1016/j.spl.2023.109836