Statistica Sinica **32** (2022), 539-545
doi:https://doi.org/10.5705/ss.202021.0191

AN EQUATION FOR THE IDENTIFICATION OF AVERAGE CAUSAL EFFECT IN NONLINEAR MODELS

Wing Hung Wong

Stanford University

Abstract: When the causal relationship between X and Y is specified by a structural equation, the average causal effect of X on Y is the population average rate of change of Y with respect to changes in X, when all other variables are kept fixed. This parameter is not identifiable from the distribution of (X, Y). We give conditions under which the average causal effect is identified as the solution of an integral equation based on the distributions of (X, Z) and (Y, Z), where Z is an instrumental variable.

Key words and phrases: Causal inference, instrumental variable.

1. Introduction

Suppose the causal relation between two real-valued random variables X and Y is specified by a structural equation $Y = f(X, U)$, where U represents all other variables that may also affect Y. We assume $f(x, U)$ is smooth in x, and write $Y(x) = f(x, U), Y^{(i)}(x) = \partial^i f(x, U)/(\partial x^i), i = 1, 2$. Then $Y^{(1)}(x)$, which tell us how Y will change when X varies around the value x, can be regarded as the causal effect of X on Y when $X = x$. This effect can be different for different subject (or sampling unit) in the population. In this paper it is assumed that we can observe X, Y but not U, the form of $f()$ is unknown, and we are interested in the estimation of the average causal effect (ACE) which is defined as the function $\theta(x) = E(Y^{(1)}(x))$. ACE is a natural generalization of $ATE = E(Y(1) - Y(0))$ when X is a binary variable indicating which of two treatments were received. ATE stands for average treatment effect, which is a parameter of central interest in the potential outcome framework for causal inference (Rubin (1974)). Since $Y(x)$ and $Y^{(i)}(x)$ are counterfactual variables (i.e. potential outcomes) that are needed in the formulation of causal relations but are not directly observable, $\theta(x)$ is not identifiable from the distribution (X, Y) alone. The method of instrumental variable attempts to identify $\theta(x)$ from the two distributions (X, Z) and (Y, Z) where the instrumental variable Z can affect X

Corresponding author: Wing Hung Wong, Departments of Statistics and Biomedical Data Science, Stanford University, Stanford, CA 94305, USA. E-mail: whwong@stanford.edu.

through another equation $X = g(Z, V)$. However, identifiability results for causal parameters typically requires monotonicity assumptions on certain arguments of the structural equations (Imbens and Angrist (1994); Angrist, Imbens and Rubin (1996); Chernozhukov, Imbens and Newey (2007); Imbens and Newey (2009); Chen et al. (2014); Torgovitsky (2015); Kennedy, Lorch and Small (2019), see Wong (2021) for further review). Importantly, the causal parameters identified under those conditions were defined as averages of counterfactuals over certain subpopulations rather than as the unrestricted average over the whole population. Since the unrestricted population average is often also of interest (e.g. when we want to know the effect of an intervention for society at large), it is useful to supplement the existing results by developing methods to identify the unrestricted average causal effect.

We consider the following nonlinear, nonparametric causal model

(1) $Y = f(X, U), Y \in R, X \in R, U \in R^p$, f is bounded and smooth in x.

(2) $X = g(Z, V), Z \in R^q, V \in R^r$.

(3) $\sup_{x,z} p_z(x) < \infty$ where $p_z()$ denotes the density function of $X(z)$.

(4) Z is independent of (U, V).

In (1), the condition that f is bounded and smooth in x means that $\sup_u |f(x, u)| < \infty$ and $\sup_u |\partial^i f(x, u)/\partial x^i| < m(x)$ for $i = 1, 2$, where $m()$ is a bounded and integrable function. Then, when $x \to \infty$, we have $Y(\infty) = \lim Y(x)$ exists and $E(Y(x)) \to E(Y(\infty))$. Similarly for $Y(-\infty)$. Also, $\theta(x) = E(Y^{(1)}(x))$ is a differentiable function and $\lim \theta(x) = 0$ as $x \to \pm\infty$. For nonlinear f and g, the independence condition (4) is not sufficient for the identification of $\theta(x)$ from the distribution of (X, Y, Z). Under the condition that changes in Y caused by varying X is uncorrelated to changes in X caused by varying Z, conditional on $Z = z$, Wong (2021) showed that the distributions (X, Z) and (Y, Z) identify a related function $\psi(z) = E(Y^{(1)}(X)|Z = z)$. That paper also demonstrated by examples that sometimes the function $\theta(x)$ can be recovered from the function $\psi(z)$, but did not provide results on the direct identification of $\theta(x)$. To fill this gap, in this paper we derive an integral equation that can be used to identify $\theta(x)$ from the distributions of (X, Z) and (Y, Z).

2. Result

To formulate our main result, consider the following conditions

(5) $I(X(z) \leq x)$ is uncorrelated with $Y^{(1)}(x)$, for all x, z.

(6) The set of distributions of $X|Z = z$, induced by varying z, is a complete set.

Theorem 1. *If* (1)–(6) *hold and* z_0 *is a fixed value, then* θ *is identifiable via the integral equation*

$$\int K(z,x)\theta(x)dx = \mu(z) - \mu(z_0) \tag{2.1}$$

where $K(z,x) = P(X \le x|Z = z_0) - P(X \le x|Z = z), \mu(z) = E(Y|Z = z)$.

Proof.

$$\begin{aligned}\mu(z) &= E(Y|Z=z) = E(f(X,U)|Z=z) = E(f(g(z,V),U)|Z=z)\\ &= E(Y(X(z))) = E\int \delta(x - X(z))Y(x)dx\end{aligned} \tag{2.2}$$

Before the formal proof we first provide a heuristic derivation. Suppose It is valid to apply integration by part to (2.2) where the delta function $\delta(t)$ is regarded as the derivative of the step function $D(t) = I(t \ge 0)$, then (2.1) follows because

$$\mu(z) = E\left(Y(\infty) - \int I(X(z) \le x)Y^{(1)}(x)dx\right) = EY(\infty) - \int P(X(z) \le x)\theta(x)dx.$$

To make this rigorous, replace $\delta()$ in (2.2) by the $N(0,\sigma^2)$ density $\varphi_\sigma()$, and define

$$\mu_\sigma(z) = E\int \varphi_\sigma(x - X(z))Y(x)dx \tag{2.3}$$

Since $Y(x) = Y(X(z)) + Y^{(1)}(X(z))(x - X(z)) + (1/2)Y^{(2)}(X(W))(x - X(z))^2$ where W is an intermediate variable lying between x and $X(z)$, we have

$$\mu_\sigma(z) = EY(X(z)) + E\left[\frac{1}{2}Y^{(2)}(X(W))\int \varphi_\sigma(x - X(z))(x - X(z))^2dx\right].$$

Thus,

$$|\mu_\sigma(z) - \mu(z)| \le \frac{\sigma^2}{2}\sup_x m(x) \le c\sigma^2 \ \text{ for some constant } c. \tag{2.4}$$

Next, we claim that there exist a constant $c > 0$, so that

$$\left|E\left(\Phi\left(\frac{x - X(z)}{\sigma}\right)Y^{(1)}(x)\right) - P(X(z) \le x)\theta(x)\right| \le cm(x)\sqrt{\sigma} \ \text{ for all small } \sigma. \tag{2.5}$$

Assuming (2.5) is true, we now analyze the integral in (2.3). Using integration

by part, we have

$$\mu_\sigma(z) = E\left[Y(\infty) - \int \left(\Phi\left(\frac{x - X(z)}{\sigma}\right)Y^{(1)}(x)\right)dx\right]$$
$$= E(Y(\infty)) - \int P(X(z) \le x)\theta(x)dx + r(z,\sigma)$$

where for some constant c, $|r(z,\sigma)| \le c\sqrt{\sigma}$ for all small σ. Thus,

$$\left|(\mu_\sigma(z) - \mu_\sigma(z_0)) - \int [P(X(z_0) \le x) - P(X(z) \le x)]\theta(x)dx\right| \le 2c\sqrt{\sigma}. \quad (2.6)$$

Taking the limit of (2.4) and (2.6) as $\sigma \to 0$, we have

$$\mu(z) - \mu(z_0) = \lim_{\sigma\to 0}(\mu_\sigma(z) - \mu_\sigma(z_0)) = \int [P(X(z_0) \le x) - P(X(z) \le x)]\theta(x)dx.$$

The desired equation (2.1) follows because $P(X(z) \le x) = P(g(z,V) \le x) = P(g(z,V) \le x|Z = z) = P(g(Z,V) \le x|Z = z) = P(X \le x|Z = z)$. To prove the claim (2.5),

$$\left|E\left(\Phi\left(\frac{x - X(z)}{\sigma}\right)Y^{(1)}(x)\right) - P(X(z) \le x)\theta(x)\right|$$
$$= \left|E\left(\Phi\left(\frac{x - X(z)}{\sigma}\right)Y^{(1)}(x)\right) - E(I(X(z) \le x))E(Y^{(1)}(x))\right|$$
$$= \left|E\left(\Phi\left(\frac{x - X(z)}{\sigma}\right)Y^{(1)}(x)\right) - E(I(X(z) \le x)Y^{(1)}(x))\right| \text{ (by condition (5))}$$
$$\le m(x)E\left|\Phi\left(\frac{x - X(z)}{\sigma}\right) - I(X(z) \le x)\right|$$
$$\le m(x)\left[\Phi\left(-\frac{1}{\sqrt{\sigma}}\right) + 4(\sup_{x,z} p_z(x))\sqrt{\sigma}\right] \quad (2.7)$$

The last inequality (2.7) holds because $|\Phi((x - X(z))/\sigma) - I(X(z) \le x)|$ is bounded by 2 on $A(\sigma)$ and by $\Phi(-1/\sqrt{\sigma})$ on $A(\sigma)^c$, where $A(\sigma)$ is the event $\{|X(z) - x| \le \sqrt{\sigma}\}$. Finally (2.5) follows from (2.7) because of the exponentially decreasing tail of the normal distribution. Since both $K(z,x)$ and $\mu(z)$ in the integral equation (2.1) are determined by the distributions of (X,Z) and (Y,Z), it follows that θ is also determined if the solution to (2.1) is unique. To establish uniqueness, let a be a fixed constant, and define for any $\theta()$, its anti-derivative $\lambda(x) = a - \int_x^\infty \theta(t)dt$. Suppose θ_1 and θ_2 are two solutions to (2.1) and λ_1 and

λ_2 are the corresponding anti-derivatives, then

$$E(\lambda_1(X) - \lambda_2(X)|Z = z) = \int p_{X|Z}(x|z)(\lambda_1 - \lambda_2)(x)dx$$
$$= -\int P(X \leq x|Z = z)(\theta_1 - \theta_2)(x)dx = -\int P(X \leq x|Z = z_0)(\theta_1 - \theta_2)(x)dx$$

Since the last expression does not depend on z, Condition (6) implies $\lambda_1 = \lambda_2$, and therefore $\theta_1 = \theta_2$.

3. Discussion

Of the 6 conditions in the theorem, the first 3 are needed just to set up the model and are not restrictive. On the other hand, conditions (4), (5), (6) each represents a significant constraint on the model. Condition (4) says that Z is independent of all other causal variables that affect X and Y. Together with (1) and (2), this means that the only way Z can affect Y causally is indirectly through its effect on X. This is a natural condition on an instrumental variable. Condition (6) implies that the family of conditional distributions $P(X|Z = z)$ as z varies, is a large family. This means that Z has non-trivial relationship with X in the sense that varying the value of z leads to rich changes in the distribution of X. This is also a reasonable condition on an instrumental variable. This type of completeness condition was first introduced into causal inference by Newey and Powell (2003). Finally, condition (5) requires the causal effect $Y^{(1)}(x) = (\partial f/\partial x)(x, U)$ to be uncorrelated to $I(X(z) \leq x) = I(g(z, V) \leq x)$, which is a strong condition. However, even in the simplest case when both X and Z are binary variables, it is not possible to identify the average treatment effect (analog of θ in that case) from the distribution (X, Z) and (Y, Z) without similarly strong conditions (see discussion in Angrist, Imbens and Rubin (1996)). In the general context of (1)–(4), we are not aware of alternative conditions that can be used to relate $\mu(z)$ to $\theta(x)$. The following example illustrates the use of our result in a nonlinear, nonparametric model that allows i) unobserved confounders and ii) heterogeneity in the causal effect of X on Y.

Example 1. Suppose $Y = h(X, U_1) + U_2, X = g(Z, V)$, where $h()$ is a smooth and bounded function in x. If U_1 is independent of V, then condition (5) is satisfied. Note that since no restriction is imposed on the joint distribution of U_2 and V, they may include unobserved confounders that affect both X and Y. Also, the completeness condition (6) is not too restrictive. For example, (6) holds in the following cases (a) $g(z, v) = s(z + v)$ where $s()$ is an invertible function

and V is a continuous random variable, (b) $g(z, v) = 1 + v_1 z + v_2 z^2, V_1$ and V_2 are independent random variables.

From the proof of the theorem, it is clear that if condition (5) holds only for some values of z and z_0, then equation (2.1) will hold for those z and z_0. If we are willing to make some modeling assumptions on $\theta()$, say $\theta(x) = \Sigma_1^k \alpha_i b_i(x)$ where $b_i(), i = 1, \ldots, k$ are fixed functions, then we may weaken condition (5) by requiring it to hold only for a finite subset of values for z and then use the corresponding finite set of equations to identify the parameters $\alpha_i, i = 1, \ldots, k$. Finally, we note that above proof of the theorem follows the way we discovered the integral equation originally, namely, start with the expression for $E(Y|Z = z)$, replace the delta function in the expression by the normal kernel and then integrate by part to obtain an expression involving $\theta()$. Weijie Su (personal communication) suggests a second proof, which starts from the given $K(z, x)$ and then shows that the integral in (2.1) gives rise to $\mu(z) - \mu(0)$. His proof has the advantage that it does not require the existence of bounded second derivatives.

Acknowledgments

The author thanks Xiaohong Chen, Peng Ding, Dylan Small, and Weijie Su for helpful comments. This work was supported by NSF grants DMS1811920 and DMS1952386.

References

Angrist, J. D., Imbens, G. W. and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association* **91**, 444–455.

Chen, X., Chernozhukov, V., Lee, S. and Newey, W. K. (2014). Local identification of nonparametric and semiparametric models. *Econometrica* **82**, 785–809.

Chernozhukov, V., Imbens, G. W. and Newey, W. K. (2007). Instrumental variable estimation of nonseparable models. *Journal of Econometrics* **139**, 4–14.

Imbens, G. and Angrist, J. (1994). Identification and estimation of local average treatment effects. *Econometrica* **62**, 467–475.

Imbens, G. W. and Newey, W. K. (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica* **77**, 1481–1512.

Kennedy, E. H., Lorch, S. and Small, D. S. (2019). Robust causal inference with continuous instruments using the local instrumental variable curve. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **81**, 121–143.

Newey, W. K. and Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica* **71**, 1565–1578.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66**, 688–701.

Torgovitsky, A. (2015). Identification of nonseparable models using instruments with small sup-

port. *Econometrica* **83**, 1185–1197.

Wong, W. H. (2021). A calculus for causal inference with instrumental variables. *arXiv preprint arXiv:2104.10633*.

Wing Hung Wong

Departments of Statistics and Biomedical Data Science, Stanford University 390 Jane Stanford Way, Stanford, CA 94305, USA.

E-mail: whwong@stanford.edu

(Received May 2021; accepted August 2021)