Institute for Economic Studies, Keio University

Keio-IES Discussion Paper Series

Identification of heterogeneous treatment effects as a function of potential untreated outcome under the nonignorable assignment condition

星野崇宏、高畑圭祐

2018 年 4 月 9 日

DP2018-005

https://ies.keio.ac.jp/publications/9396/

Institute for Economic Studies, Keio University
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
ies-office@adst.keio.ac.jp

9 April, 2018

Identification of heterogeneous treatment effects as a function of potential untreated outcome under the nonignorable assignment condition

星野崇宏、高畑圭祐

IES Keio DP2018-005
2018 年 4 月 9 日
JEL Classification: C13; C31; C83
キーワード: 無視できない欠測;因果効果;識別性

【要旨】

We provide sufficient conditions for the identification of hetero- geneous treatment effects (HTE), in which the missing mechanism is nonignorable, when the information on the marginal distribution of untreated outcome is available. It is also shown that, under such a situ- ation, the same result holds for the identification of average treatment effects (ATE). Exposing certain additivity on the regression function of the assignment probability, we reduce the identication of HTE to the uniqueness of a solution of some integral equation, and discuss it borrowing the idea from the literature on statistical inverse prob- lems. Our result contributes to theoretical understandings in causal inference with heterogeneity and also the relaxation of the conditional independence assumption in statistical data fusion or statistical data combination.

星野崇宏
慶應義塾大学経済学部
〒108-8345
東京都港区三田2ー15ー45
hoshino@econ.keio.ac.jp

高畑圭祐
慶應義塾大学大学院 経済学研究科
〒108-8345
東京都港区三田2ー15ー45
kei_t_12@keio.jp

謝辞：本研究は科研費（１８Ｈ０３２０９）による研究成果である。ここに記して謝意を表したい。

Identification of heterogeneous treatment effects as a function of potential untreated outcome under the nonignorable assignment condition

Keisuke Takahata and Takahiro Hoshino

Abstract

We provide sufficient conditions for the identification of heterogeneous treatment effects (HTE), in which the missing mechanism is nonignorable, when the information on the marginal distribution of untreated outcome is available. It is also shown that, under such a situation, the same result holds for the identification of average treatment effects (ATE). Exposing certain additivity on the regression function of the assignment probability, we reduce the identification of HTE to the uniqueness of a solution of some integral equation, and discuss it borrowing the idea from the literature on statistical inverse problems. Our result contributes to theoretical understandings in causal inference with heterogeneity and also the relaxation of the conditional independence assumption in statistical data fusion or statistical data combination.

Keyword: *nonignorable missing; identifiability; auxiliary information; causal inference; statistical data fusion; integral equation*

1 Introduction

In observational studies, a treatment effect of interest is generally defined as average treatment effects (ATE) or average treatment effects on the treated (ATT). The strong ignorability condition, which requires an assignment to be independent of potential outcomes given covariates, is well-known to play a significant role in the identification of those quantities (Rosenbaum and Rubin, 1983). On the other hand, at times, researchers want to know more individualized or heterogeneous causal effects, while, as its name suggests, ATE and ATT are averaged effects over the population or a subset of the population. Estimation of such effects has been paid attention to in recent years (e.g., Wager and Athey, 2017), particularly in marketing and medicine where personalized treatments are known to be effective, but theoretical aspects have not been necessarily studied sufficiently. Considering this purpose, we discuss the identification of the following *heterogeneous treatment effects* (HTE), which we define in this paper as

$$\text{HTE}(y_0) \equiv E[y_1 - y_0|y_0] = E_{x|y_0}[E[y_1|y_0, x]] - y_0, \tag{1}$$

where $y_1 \in \mathbb{R}$ and $y_0 \in \mathbb{R}$ are the outcome variables when assigned to the treatment group and the control group respectively, and $x \in \mathbb{R}^d$ is a d-dimensional covariate vector. $E_{x|y_0}[\cdot]$ denotes the expectation over x given y_0. HTE implies how much effect people whose outcome is y_0 under the untreated condition would get if assigned to the treatment. Therefore, HTE is a function of y_0.

Although HTE may have an implication for researchers' interest, its identification is not trivial owing to the dependence of unobserved variable: we need to identify the density of y_1 given y_0 and x, $p(y_1|y_0, x)$, as seen in equation (1), but y_1 and y_0 are never observed simultaneously. Therefore, additional conditions are needed for the identification. In this paper, we obtain relaxing the strong ignorability condition as

$$p(z|y_1, y_0, x) = p(z|y_0, x), \tag{2}$$

where $z \in \{0, 1\}$ is an assignment indicator, which is $z = 1$ when assigned to the treatment group. We refer to this assumption as *weak ignorability*. This assumption is justifiable for the following two reasons. First, it is always weaker than the strong ignorability assumption. Second, since an assignment, z, precedes the outcome in causal inference, it is natural to assume that an assignment to the treatment should be influenced by the default value of the outcome, y_0, rather than by the outcome with some special treatment, y_1. It would not be straightforward to observe how weak ignorability works in the identification of HTE, but the details will be described in section 2.

However, weak ignorability still requires us to consider dependence on unobserved outcome y_0 in the treatment group, that is, the missing mechanism is nonignorable (or NMAR, Little and Rubin, 2002). Under nonignorable missingness, full parametric assumptions, which cannot be tested by original data alone, are generally needed for identification (e.g., Miao et al., 2017; Cui et al., 2017). One of the most popular approaches to deal with this issue is to use auxiliary information (Hirano et al., 2001; Nevo, 2003; Deng et al., 2013; Si et al., 2014; Chen et al., 2017). Following this approach, we provide sufficient conditions for the identification of HTE with auxiliary information under weak ignorability. We assume that the marginal distribution of an outcome under the untreated condition, $p(y_0)$, is available. As a treatment is generally conducted on a small subset of the target population and y_0 is the outcome when not assigned to such treatment, we may use information on $p(y_0)$ outside the experiment. For example, survival time distribution in the population is available in medical research, or income distribution is estimated by using the census in economics.

Hirano et al. (2001) considered a situation where there is nonignorable attrition in a two-period panel, while refreshment samples, which are new additional units randomly sampled from the target population, are available. They provided sufficient conditions for identification in this setting. In this paper, we reduce the identifiability of HTE to the uniqueness of a solution of some integral equation. The integral equation to solve here has the same structure as that of nonparametric instrumental variable models; the uniqueness of the integral equation is then discussed based on Newey and Powell (2003), which characterized uniqueness of the integral equation as completeness of certain conditional distribution.

Based on these results, we show that, with the information on $p(y_0)$, it is sufficient for the identification of HTE that the extended version of the propensity score (described later) (i) is specified by the logistic regression, (ii) the regression function has no interaction term between y_0 and x, and (iii) its part relating to y_0 is linear in parameters (but note that it can be a nonlinear function of y_0). It is also shown that the same conditions are sufficient for the identification of ATE under weak ignorability, and that only the information on $p(y_0)$ is sufficient for ATT. In addition, it is notable that our result ensures point identification, while there are several literature providing partial identification results on statistical data combination, where outcome variables and covariates are separately observed (Manski, 2000; Fan et al., 2014; Fan et al., 2017).

2 Identification of ATT, ATE, and HTE

Let us assume that the marginal distribution of the untreated outcome, $p(y_0)$, is known. ATE and ATT can be formulated as

$$\text{ATE} \equiv E[y_1 - y_0] = E_{y_0,x}[E[y_1|y_0, x]] - E[y_0], \tag{3}$$

$$\text{ATT} \equiv E[y_1 - y_0|z = 1] = E[y_1|z = 1] - E[y_0|z = 1]. \tag{4}$$

Note that we can consistently estimate $E[y_1|z = 1]$ through observed data and $E[y_0]$ by the assumption. Besides, as $p(y_0|z = 1)$ can be calculated by

$$p(y_0|z = 1) = \frac{p(y_0) - p(y_0|z = 0)\Pr(z = 0)}{\Pr(z = 1)}, \tag{5}$$

the identifiability of ATT is trivial. Therefore, it suffices to provide conditions for the identification of $p(y_0, x)$ and $p(y_1|y_0, x)$ for ATE, and this is clearly sufficient for HTE.

First, we discuss the identification of $p(y_0, x)$. By Bayes' rule, $p(y_0, x)$ can be written as

$$p(y_0, x) = \frac{p(y_0, x|z = 0)\Pr(z = 0)}{\Pr(z = 0|y_0, x)} = \frac{p(y_0, x|z = 0)\Pr(z = 0)}{1 - \Pr(z = 1|y_0, x)}. \tag{6}$$

According to Hirano et al. (2001)'s main theorem, when $p(y_0)$ is known, $p(z = 1|y_0, x)$ is identified in the following form:

$$\Pr(z = 1|y_0, x) = 1 - g(k_0 + k_{y_0}(y_0) + k_x(x)), \tag{7}$$

where g is a known function that is differentiable, strictly increasing with $\lim_{x\to-\infty} g(x) = 0$ and $\lim_{x\to\infty} g(x) = 1$, and $k_{y_0}(\cdot)$ and $k_x(\cdot)$ are a unique set of functions subject to normalizations, $k_{y_0}(0) = k_x(0) = 0$. Equation (7) can be interpreted as the extended version of the propensity score (hereafter, simply "the propensity score"), while the original version of the propensity score is generally defined as the probability of being assigned to the treatment group given only covariates (Rosenbaum and Rubin, 1983). The point is that the propensity score must be specified not including an interaction term between y_0 and x, that is, additivity has to hold in equation (7). Considering this result, it is straightforward to observe that $p(y_0, x)$ is identifiable.

Next, we discuss the identification of $p(y_1|y_0, x)$. Let us consider the following integral equation:

$$
\begin{aligned}
p(y_1|x, z = 1) &= \int p(y_1|y_0, x, z = 1)p(y_0|x, z = 1)dy_0 \\
&= \int p(y_1|y_0, x)p(y_0|x, z = 1)dy_0, \qquad (8)
\end{aligned}
$$

where the second equality holds due to weak ignorability. Note that we can consistently estimate $p(y_1|x, z = 1)$ through observed data. Moreover, by substituting $z = 0$ with $z = 1$ in the middle part of equation (6), it is easily verified that $p(y_0|x, z = 1)$ is identifiable. Hence, if there is a unique $p(y_1|y_0, x)$ that satisfies equation (8), then $p(y_1|y_0, x)$ is identifiable, which leads to the identifiability of HTE and ATE.

Equation (8) is called Fredholm integral equation of the first kind, and these types of equations are known to be ill-posed problems. In other words, additional conditions or regularizations are needed to obtain a stable and unique solution. Several problems, such as nonparametric instrumental variable models or measurement-error models, are known to reduce to solve this type of equations, and are studied in econometrics as statistical inverse problems (see for example, Carrasco et al., 2007; Horowitz, 2009). Newey and Powell (2003) characterized the uniqueness of a solution of this type of integral equation as completeness of the distribution with respect to which the expectation of the function of interest is taken. Since their pioneering work, the completeness condition has been widely used in econometrics (e.g., Hall and Horowitz, 2005; Blundell et al., 2007; Darolles et al., 2011; Horowitz, 2011; Horowitz and Lee, 2012). Following Newey and Powell's discussion, we provide sufficient conditions for the identification of HTE and ATE.

Theorem. Under weak ignorability, if (i) $p(y_0)$ is known, (ii) the propensity score is specified by the logistic regression, and (iii) $k_{y_0}(\cdot)$ in equation (7) is linear in the parameter θ_{y_0}, then HTE and ATE are identifiable.

Proof. Let $p(y_1|y_0, x)$ and $\tilde{p}(y_1|y_0, x)$ be any solutions of equation (8). By subtracting both equations with solution inserted, we obtain

$$\int (p(y_1|y_0,x) - \tilde{p}(y_1|y_0,x))p(y_0|x,z=1)dy_0$$
$$= \int h(y_0,x)p(y_0|x,z=1)dy_0 = 0, \tag{9}$$

where $h(y_0, x) = p(y_1|y_0, x) - \tilde{p}(y_1|y_0, x)$. If it is the case that $h(y_0, x)$ that solves equation (9) is always zero for all x, then $p(y_1|y_0, x)$ is equal to $\tilde{p}(y_1|y_0, x)$ and equation (8) has a unique solution. This implies that, if $p(y_0|x, z = 1)$ is complete, $p(y_1|y_0, x)$ is identifiable. Therefore, it suffice to find sufficient conditions for $p(y_0|x, z = 1)$ being complete.

Let us assume that the function g in equation (7) is the distribution function of logistic distribution. In this case, $p(y_0|x, z = 1)$ can be rewritten as

$$p(y_0|x,z=1) = \frac{p(y_0,x)\Pr(z=1|y_0,x)}{p(x,z=1)}$$
$$= \exp(k_0 + k_{y_0}(y_0) + k_x(x))\frac{p(y_0,x|z=0)\Pr(z=0)}{p(x,z=1)}. \tag{10}$$

Pugging equation (10) into equation (9) yields

$$\int \exp(k_{y_0}(y_0))m(y_0,x)dy_0 = 0 \tag{11}$$

where

$$m(y_0,x) = h(y_0,x)\exp(k_0 + k_x(x))\frac{p(y_0,x|z=0)\Pr(z=0)}{p(x,z=1)}. \tag{12}$$

Here, let us introduce the additional assumption. Denote by θ_{y_0} the parameter vector of $k_{y_0}(y_0)$: $k_{y_0}(y_0) = k_{y_0}(y_0; \theta_{y_0})$. We assume that $k_{y_0}(y_0; \theta_{y_0})$ is linear in the parameter θ_{y_0}:

$$k_{y_0}(y_0;\theta_{y_0}) = \sum_{j=1}^{p} \theta_{y_0 j} T_j(y_0), \tag{13}$$

where $T_j(y_0)$ is any statistics of y_0. Substituting $k_{y_0}(y_0)$ in equation (11) with equation (13), we obtain

$$\int \exp\left(\sum_{j=1}^{p} \theta_{y_0 j} T_j(y_0)\right) m(y_0,x)dy_0 = 0. \tag{14}$$

Equation (14) has the same structure as the discussion about completeness of exponential family. Therefore, considering Theorem 1 (Lehman, 1986, p. 142), it leads to $m(y_0, x) = 0$ for all x. As each value on the right-hand side of (12) is strictly positive except $h(y_0, x)$, it follows that $h(y_0, x) = 0$, and equation (8) has a unique solution. □

The assumption on $k_{y_0}(\cdot)$ does not require $k_{y_0}(\cdot)$ to be a linear function of y_0, but to be a linear function of the parameter θ_{y_0}. Therefore, $k_{y_0}(\cdot)$ can include, for example, a polynomial of y_0. Given that the true function can be approximated by the finite Taylor series, this assumption is not that strong. On the other hand, the additivity condition between y_0 and x may be strong. As long as Hirano et al.'s result is followed, this condition must be satisfied. However, if we can make use of the information on $p(y_0, x)$ from other sources, this can be relaxed and we can identify an interaction term, $k_{y_0,x}(\cdot, \cdot)$.

3 Discussion

The completeness condition, as used in the proof of the theorem, has been paid attention to in recent years since Newey and Powell (2003). Although it is widely applicable, Canay et al. (2013) showed that the completeness condition cannot be tested by observed data. This implies that "for every complete distribution, there exists an incomplete distribution which is arbitrarily close to it (Freyberger, 2017)." Canay et al. (2013) argued that their result did not suggest to avoid using the completeness condition but to justify it with alternative arguments. On the other hand, there are several papers which provide sufficient conditions as alternatives for the completeness condition, which may be testable (Newey and Powell, 2003; D'Haultfoeuille, 2011; Hu and Shiu, 2016). Our assumption that the function g is the distribution function of the logistic distribution can be related to the latter approach. Newey and Powell (2003) provided the sufficient condition that the certain conditional distribution corresponding to $p(y_0|x, z = 1)$ in our model is of the exponential family, but the specification by the logistic regression may be rather weaker.

Our result can be related to the partial identification literature on statistical data fusion or statistical data combination (Ridder and Moffitt, 2007). Fan et al. (2014) considered a situation where outcome variables and covariates are separately observed and partial identification results are derived.

Although they assumed strong ignorability, that is, the missing mechanism is ignorable, there was no sample observed as a set of the outcome and covariates. On the other hand, we assumed that the outcome and covariates were observed simultaneously for each group, hence their setting is more general than ours in this sense. However, we considered the nonignorable missing mechanism (weak ignorability) and provided point identification results by using auxiliary information. Therefore, which approach is more useful may depend on a situation.

Moreover, our approach can be directly extended to statistical data fusion to relax the conditional independence assumption. Suppose that dataset A contains $y_0 \in \mathbb{R}$ and covariates $x \in \mathbb{R}^d$, dataset B contains $y_1 \in \mathbb{R}$ and x, and x are common in both the datasets. The aim is to impute $y_1(y_0)$ in dataset A (B) given $y_0(y_1)$ and x (respectively). In other words, $p(y_1|y_0, x)$ and $p(y_0|y_1, x)$ are to be estimated. However, this implies that the missing mechanism is nonignorable. Therefore, the conditional independence, $p(y_1, y_0|x) = p(y_1|x)p(y_0|x)$, is generally assumed and missing values are imputed by $p(y_1|x)$ and $p(y_0|x)$. Although this assumption makes the missing mechanism ignorable and tractable, it is rather arbitrary, particularly when covariates do not have sufficient information. In data fusion, it can be the case that the marginal distribution of each target variable is known from other sources but the joint distribution $p(y_0, y_1)$ is unknown. In this case, our result implies that $p(y_1|y_0, x)$ and $p(y_0|y_1, x)$ can be consistently estimated with such auxiliary information, that is, we do not have to assume conditional independence. However in this case the weak ignorability condition may need some modification.

4 Conclusion

We provided sufficient conditions for the identification of HTE with information on the marginal distribution of untreated outcome under the nonignorable missing assumption, the same result for ATE, and weaker conditions for ATT. Our result contributes to the understanding of theoretical aspects of treatment effects with heterogeneity, which are focused in marketing, medicine and many other applications. It is also shown that our result may be useful for statistical data fusion in relaxing the conditional independence assumption.

As we discussed only identification in this paper, we are now planning to develop an estimation procedure. As it includes solving integral equations such as equation (8), we may borrow existing methods from inverse problems, such as sieve estimation or Tikhonov regularization. One aspect we need to consider is to incorporating the information on the marginal distribution into the optimization problem as a certain constraint and we do not consider it to be very challenging to do so.

References

[1] Blundell, R., X. Chen, and D. Kristensen (2007): "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica*, 75, 1613-1669.

[2] Canay, I. A., A. Santos, and A. M. Shaikh (2013): "On the Testability of Identification in Some Nonparametric Models With Endogeneity," *Econometrica*, 81, 2535-2559.

[3] Carrasco, M., J.P. Florens, and E. Renault (2007): "Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization," In Heckman, J. and E. Leamer (eds.), *Handbook of Econometrics*, 6B (Ch. 77). North-Holland: Elsevier Science.

[4] Chen, H., M.-H. Felt, and K.P. Huynh (2017): "Retail payment innovations and cash usage: accounting for attrition by using refreshment samples," *Journal of the Royal Statistical Society: A*, 180, 503-530.

[5] Cui, X., J. Guo, and G. Yang (2017): "On the identifiability and estimation of generalized linear models with parametric nonignorable missing data mechanism," *Computational Statistics and Data Analysis*, 107, 64-80.

[6] Darolles, S., Y. Fan, J. P. Florens, and E. Renault (2011): "Nonparametric Instrumental Regression," *Econometrica*, 79, 1541-1565.

[7] Deng, Y., D.S. Hillygus, J.P. Reiter, Y. Si, and S. Zheng (2017): "Handling Attrition in Longitudinal Studies: The Case for Refreshment Samples," *Statistical Science*, 28, 238-256.

[8] D'Haultfoeuille, X. (2011): "On the Completeness Condition in Nonparametric Instrumental Problems," *Econometric Theory*, 27, 460-471.

[9] Fan, Y., E. Guerre, and D. Zhu (2017): "Partial identification of functionals of the joint distribution of "potential outcomes"", *Journal of Econometrics*, 197, 42-59.

[10] Fan, Y., R. Sherman, and M. Shum (2014): "Identifying Treatment Effects under data combination", *Econometrica*, 82, 811-822.

[11] Freyberger, J. (2017): "On Completeness and Consistency in Nonparametric Instrumental Variable Models", *Econometrica*, 85, 1629-1644.

[12] Hall, P., and J.L. Horowitz (2005): "Nonparametric Methods for Inference in the Presence of Instrumental Variables," *The Annals of Statistics*, 33, 2904-2929.

[13] Hirano, K., G.W. Imbens, G. Ridder, and D.B. Rubin (2001): "Combining Panel Data Sets with Attrition and Refreshment Sampes," *Econometrica*, 69, 1645-1659.

[14] Horowitz, J.L. (2009): "Semiparametric and nonparametric methods in econometrics," New York: Springer-Verlag.

[15] Horowitz, J.L. (2011): "Applied nonparametric instrumental variable regression," *Econometrica*, 79, 347-394.

[16] Horowitz, J. L., and S. Lee (2012): "Uniform Confidence Bands for Functions Estimated Nonparametrically With Instrumental Variables," *Journal of Econometrics*, 168, 175-188.

[17] Hu, Y. and J.-L. Shiu (2017): "Nonparametric Identification Using Instrumental Variables: Sufficient Conditions for Completeness," *Econometric Theory*, 11-35. doi:10.1017/S0266466617000251.

[18] Lehmann, E.L. (1986): "Testing statistical hypotheses (2nd ed.)," New York: Springer.

[19] Little, R.J.A., and D.B. Rubin (2002): "Statistical analysis with missing data (2nd ed.)," New York: Wiley.

[20] Manski, C.F. (2000): "Identification Problems and Decision Under Ambiguity: Empirical Analysis of Treatment Response and Normative Analysis of Treatment Choice," *Journal of Econometrics*, 95, 415-442.

[21] Miao, W., P. Ding, and Z. Geng (2017): "Identifiability of Normal and Normal Mixture Models with Nonignorable Missing Data," *Journal of the American Statistical Association*, 111, 1673-1683.

[22] Nevo, A. (2003): "Using Weights to Adjust for Sample Selection When Auxiliary Information Is Available", *Journal of Business and Economic Statistics*, 21, 43-52.

[23] Newey, W.K. and J.L. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565-1578.

[24] Ridder, G., R. Moffitt (2007): "The econometrics of data combination," In Heckman, J. and E. Leamer (eds.), *Handbook of Econometrics*, 6B (Ch. 75). North-Holland: Elsevier Science.

[25] Rosenbaum, P.R. and D. B. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41-55.

[26] Si, Y., J.P. Reiter, and D.S. Hillygus (2015): "Semi-parametric Selection Models for Potentially Non-ignorable Attrition in Panel Studies with Refreshment Samples," *Political Analysis*, 23, 92-112.

[27] Wager, S. and S. Athey (2017): "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests," *Journal of the American Statistical Association*, DOI: 10.1080/01621459.2017.1319839.