Summary: Essays on Estimation of Causal Parameters Identified as the Ratio of Conditional Expectation Functions

Kazuhiko Shinoda

Graduate School of Economics, Keio University

This thesis consists of three essays on the estimation of causal parameters that are identified as the ratio of conditional expectation functions (CEFR), and an introduction to the problem of CEFR estimation in causal inference.

Chapter 1 Introduction

This chapter introduces key concepts and provides a brief background to the issues studied in this thesis. The potential outcome approach (Rubin, 1974; Imbens and Rubin, 2015) is arguably the most widely used framework for causal inference in the social sciences. Consider a simple case with a binary treatment, i.e. individuals may receive a treatment or not. Let Y be an outcome of interest, and D be a treatment status that takes 1 if an individual receives a treatment and 0 otherwise. For example, when we are interested in assessing the effect of job training on annual income after a year, we can set the receipt of training as D, and the annual income after a year as Y. The potential outcomes are the response that would be observed if an individual were assigned a specific treatment. Let Y_1 and Y_0 be the potential outcomes with and without a treatment, and assume that the realized outcome always satisfies $Y = DY_1 + (1 - D)Y_0$.

Using the potential outcomes, we can calculate the causal effect of a treatment on an outcome of interest as $Y_1 - Y_0$. However, this calculation is never feasible in practice as we can never observe both Y_1 and Y_0 for the same individual. This is known as *the fundamental problem of causal inference* (Holland, 1986), indicating that the true value of treatment effects is never observed. The most prevalent approach to overcome the fundamental problem of causal inference is assuming *unconfoundedness* and measuring a treatment effect by *the average treatment effect (ATE)* $E[Y_1 - Y_0]$. Unconfoundedness states that (Y_1, Y_0) and D are independent, which allows estimation of ATE from the observed variables as:

$$E[Y_1 - Y_0] = E[Y_1] - E[Y_0] = E[Y_1|D = 1] - E[Y_0|D = 0] = E[Y|D = 1] - E[Y|D = 0],$$

where the second equality holds by unconfoundedness, and the last equality follows from the assumption

$Y = DY_1 + (1 - D)Y_0$.

Unconfoundedness introduced above is very restrictive since it requires that the treatment status is independent of the potential outcomes unconditionally. Therefore, the following conditional version is more popular in the literature:

$$Y_1, Y_0 \perp\!\!\!\perp D|X, \tag{1}$$

where $A \perp\!\!\!\perp B|C$ means A and B are conditionally independent given C, and X is a vector of covariates. In the job training example, X may include variables characterizing individuals such as age, gender, education and so on. If the conditional unconfoundedness holds, we can estimate ATE as:

$$\begin{aligned} E[Y_1 - Y_0] &= E[Y_1] - E[Y_0] = E[E[Y_1|X]] - E[E[Y_0|X]] \\ &= E[E[Y|D = 1, X]] - E[E[Y_0|D = 0, X]]. \end{aligned}$$

Although ATE can be used to evaluate the size of a treatment effect averaged over the population of interest, it cannot capture the heterogeneous nature of the effect for different individuals. Quantifying such heterogeneity is essential in many research areas including precision medicine and target marketing. Among several different definitions of heterogeneous treatment effects, the most actively studied is a treatment effect conditional on covariates X, such as *the conditional average treatment effect (CATE)* $E[Y_1 - Y_0|X]$. All the subsequent chapters deal with the estimation of heterogeneous treatment effects.

There are several treatment effects that are identified as CEFR, and this thesis focuses on the estimation and inference on such causal parameters. A simple example is ratio-based treatment effects such as the odds ratio, hazard ratio and survival ratio (Dukes and Vansteelandt, 2018; Liang and Yu, 2020; Yadlowsky et al., 2021; Lee, 2022). They are defined as the ratio of the expected potential outcomes $E[Y_1|X]/E[Y_0|X]$, rather than the difference $E[Y_1|X] - E[Y_0|X]$, and the ratio is a more natural choice to measure the causal effect such as a treatment effect on the relapse rate of a specific decease, cost saving effects, the effect of dieting on weight. For example, a fixed amount of weight loss can be of different importance to individuals with different levels of baseline weight. Chapter 4 develops a novel estimation method for a ratio-based treatment effect from time-to-event data.

In some problem settings, we can estimate some difference-based treatment effects as in the form of CEFR. A typical example is LATE. Besides that, Yamane et al. (2018) showed that CATE is identified as CEFR in the data combination setting, where Y and D cannot be observed simultaneously in a single dataset. Chapter 2 extends their work to the estimation of LATE by data combination. Chapter 3 develops

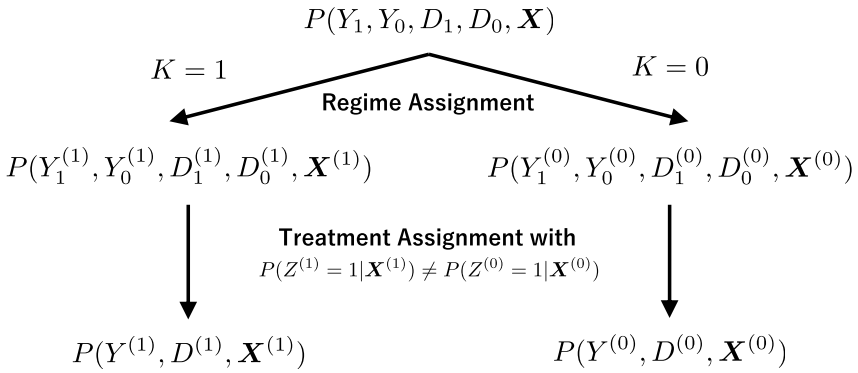


Figure 1: Overview of the Problem Setting.

the general inference framework for CEFR, which can handle both ratio-based treatment effects and data combination, among many other examples.

Chapter 2 Estimation of Local Average Treatment Effect by Data Combination

Estimating the causal effects of treatment on an outcome of interest is central to optimal decision making in many real-world problems. However, the identification and estimation of treatment effects usually rely on the untestable assumption referred to as *unconfoundedness*, namely, independence between the treatment status and potential outcomes (Imbens and Rubin, 2015). Violations of unconfoundedness may occur not only in observational studies, but also in randomized controlled trials (RCTs) when compliance with the assigned treatment is not complete. For example, even if a coupon is distributed randomly to measure its effect on sales, the probability of using the coupon is likely to depend on the unobserved nature of the individuals. Moreover, noncompliance can also occur regardless of the individual's intentions. In online advertisement placement, the probability of watching the ad depends on the bidding strategy of other companies because ads that are actually displayed are determined through the real-time-bidding even if one tries to place the ad randomly.

In such cases, it is well known that LATE can be identified and estimated using the treatment assignment as an instrumental variable and conditions milder than unconfoundedness (Imbens and Angrist, 1994; Angrist et al., 1996; Frölich, 2007). LATE is the treatment effect measured for the subpopulation of compliers, individuals who always follow the given assignment.

We suppose that we cannot observe all relevant variables in a single dataset for technical or privacy reasons. For example, in online-to-offline marketing, where treatments are implemented online and outcomes are observed offline, it is often difficult to match the records of the same individuals observed separately online and offline. Additionally, with the global anti-tracking movement gaining momentum,

it may become more difficult to combine multiple pieces of information online as well. Although causal inference by data combination has been actively studied (Ridder and Moffitt, 2007; Bareinboim and Pearl, 2016; Lee et al., 2020), LATE estimation using multiple datasets has not received much attention despite its practical importance. We extend the problem setting considered in (Yamane et al., 2018), where two different treatment regimes are available, to allow for the existence of noncompliance (Figure 1). Under the extended setting, LATE is shown to be identified as:

$$\frac{E[Y^{(1)}|X] - E[Y^{(0)}|X]}{E[D^{(1)}|X] - E[D^{(0)}|X]},$$

where Y is the outcome of interest, D is the treatment status, X is a vector of the baseline covariates, and a superscript represents the treatment regime.

For the estimation, we show that the direct estimation method originally developed for CATE under the complete compliance (Yamane et al., 2018) can be applied to the LATE estimation in our setting. However, their method has a practical issue in that model selection and hyperparameter tuning can be unstable owing to its minimax objective formulation. We then propose a weighted least squares estimator to avoid the minimax objective, and improve the stability in practice. Unlike the inverse probability weighted (IPW) estimator (Wooldridge, 2002, 2007; Seaman and White, 2013), which is often employed to estimate treatment effects, the proposed estimator directly uses the estimated propensity-score-difference as a weight without inversion. Therefore, our method can also avoid the common issue in the IPW methods, that is, high variance at points with a propensity score extremely close to zero.

The contributions of this study lie in the following three parts. First, we show that LATE is identified even when an outcome and treatment status cannot be observed simultaneously in a single dataset, and the treatment assignment is completely missing. Second, we find that the positivity assumption, which is necessary in the standard setting with one regime, can be omitted in our setting. We show this relaxation of the conditions further facilitates data collection. Third, we develop a novel estimator that enables simpler model selection while maintaining the essence of direct estimation as much as possible.

Our method displayed stable performance in the synthetic experiments. We also apply the proposed method to estimate the effect of job training on the total earnings over 30 months after a random assignment using the dataset of the National Job Training Partnership Act study. The result was consistent with the previous studies. Although the promising performance of our method has been verified, there are sometimes concerns about the homogeneity of the populations from which the separate samples come. This issue is addressed in the next chapter.

Chapter 3 Orthogonal Series Estimation for the Ratio of Conditional Expectation Functions

In various fields of data science, researchers often face problems of estimating CEFR. Although CEFR appears not only in causal inference studies, there are several important examples of CEFR in the treatment effect estimation literature. When an outcome of interest is the relapse rate of a specific disease, it is more natural to consider the ratio of the conditional expectation of potential outcomes $E[Y_1|X]/E[Y_0|X]$, rather than the difference $E[Y_1|X] - E[Y_0|X]$, as a measure of causal effects of a medical treatment, where Y_1 and Y_0 are the potential outcome with and without a treatment, respectively, and X is a vector of baseline covariates. Likewise, ratio-based treatment effects such as the odds ratio and hazard ratio have been widely used especially in clinical settings. Furthermore, in a data combination setting where an outcome and a treatment status are only separately observed, both CATE and LATE are identified in the form of CEFR (Yamane et al., 2018; Shinoda and Hoshino, 2022), while these effects are defined as the difference of the potential outcomes.

In this chapter, I start by developing a novel series estimator for CEFR in a very simple setting without selection bias in observed data. This series estimator is itself useful in estimating treatment effects when data can be collected completely at random from the population of interest, but such randomized data are often not available in practice. Technically, when there is selection bias in collected data, we need to estimate potentially infinite-dimensional nuisance parameters to adjust for the bias, but these parameters may be hard to estimate with a "sufficiently high quality" in observational studies on complex systems since they can be very high-dimensional and/or highly nonlinear. The highly complex nuisance parameters do not satisfy the traditional assumptions that limit the complexity of a function class, and therefore the resulting semiparametric estimator fails to be $\sqrt{N}$-consistent. I employ the debiased machine learning (DML), a set of techniques to enable the use of flexible machine learning (ML) methods for the nuisance estimation, to develop a simple and general framework for constructing a high quality estimator for CEFR even in the presence of selection bias in observed data.

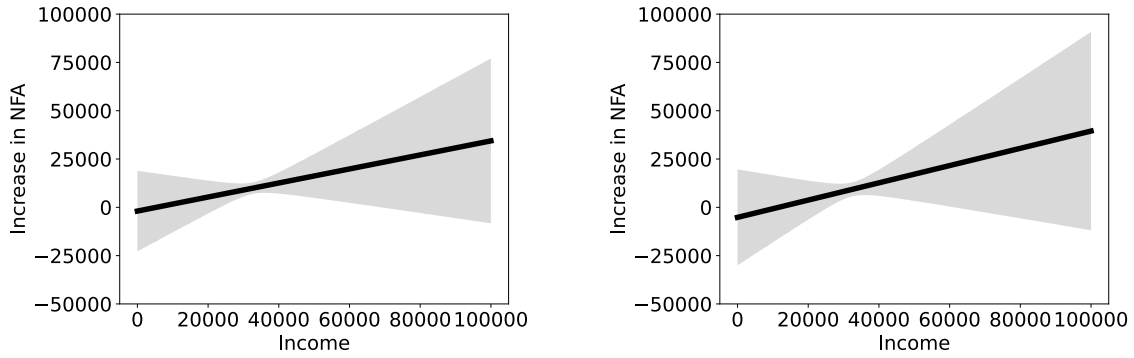
To the best of my knowledge, the previous works on the estimation of the ratio-based treatment effects all impose assumptions on the functional form somewhere in the model (Dukes and Vansteelandt, 2018; Liang and Yu, 2020; Yadlowsky et al., 2021; Lee, 2022). For example, Liang and Yu (2020) supposes that the ratio-based treatment effects can be expressed as the monotone single index model, and Yadlowsky et al. (2021) imposes a stronger condition that the monotone link function is an exponential function. On the other hand, this study considers a fully nonparametric model for treatment effects, imposing

no assumptions on the functional form of CEFR. Furthermore, Yamane et al. (2018) and Shinoda and Hoshino (2022) show that difference-based CATE and LATE are identified in the form of CEFR in the data combination setting where we cannot observe an outcome and a treatment status simultaneously in a single dataset, and develop estimation methods for CEFR. This study accommodates the generalized version of their problem settings, where each separate dataset contains selection bias. Their estimators cannot handle the selection bias without introducing additional nuisance parameters, but their methods are not orthogonal to the nuisance parameters.

The major contribution of this study is the development of the novel inference framework for CEFR with theoretical guarantees. I derive the general asymptotic results for estimation and uniform inference on the best linear approximation to the target CEFR under the proposed framework, including the validity of the Gaussian bootstrap. It is worth noting that we do not have to add stronger regularity conditions than the assumptions previously known in the literature to establish the theoretical results in this chapter. In addition to the general results, I provide a set of low-level sufficient conditions to apply the proposed framework to several specific settings. Besides the asymptotic analysis, I conduct numerical simulations to evaluate the performance of the proposed method on finite samples.

I also apply the proposed framework to estimate the causal effect of participation in the 401(k) program on household net financial assets. I conduct analyses based on the usual one-sample LATE estimation and two-sample LATE estimation, which is similar to the data combination setting considered in Chapter 2. The parameter of interest is LATE as a function of income. Figure 2(a) and (b) illustrate the estimated LATE function of income and its 95% uniform confidence band constructed by one-sample estimation and two-sample estimation, respectively. The point estimate of LATE for households with annual income \$50000 is \$16225 in one-sample estimation and \$17141 in two-sample estimation, which are consistent with the analysis in Ogburn et al. (2015) whose estimate is \$14910.

The present study also has some limitations. By construction of the proposed method, we can perform model selection based on cross validation (CV) only when the denominator function of CEFR is strictly positive almost surely. In several practical settings, the positivity of the denominator function is guaranteed, and it is shown in the simulation study that CV for the proposed estimator works well. However, there are also situations where the denominator function can take both positive and negative values, such as LATE estimation. Therefore, a model selection method for the general situation is key to increasing the practicality of the proposed framework. Despite its practical importance, little attention has been paid to model selection in the treatment effects estimation (Schuler et al., 2018; Caron et al., 2020). To the best of my knowledge, there exist only a few attempts in the literature to develop a flexible



(a) One-Sample Estimation.

(b) Two-Sample Estimation.

Figure 2: Estimated LATE of 401(k) on Net Financial Assets Conditioned on Income Level and 95% Uniform Confidence Band.

method for selecting the treatment effect model (Brookhart and van der Laan, 2006; Rolling and Yang, 2014; Saito and Yasui, 2020). Although we may be able to extend the ideas of these previous studies for the CEFR problems, further research on model selection is essential to enhance the feasibility of causal inference methods.

One of the drawbacks of the proposed framework is the large variability found in the simulation. It may be due to the structure of the orthogonal signals, in which the inverse of the estimated propensity score is used. Inverse probability weighting (IPW) is known to suffer from unstable estimates especially when the propensity score is close to zero. This problem is especially acute in the data combination settings including two-sample estimation of LATE, where we have to perform four-class or eight-class classification to estimate propensity scores. Although we can increase stability by trimming small probabilities, determining the optimal threshold is nontrivial, and trimming can cause additional bias. Recently developed *automatic debiased machine learning (Auto-DML)* (Chernozhukov et al., 2022b) can be an effective solution to the problem because it avoids the estimation of propensity scores. Auto-DML directly estimates the Riesz representer of the orthogonal signals rather than constructing it with the inverse of the estimated propensity score. Much smaller variance of Auto-DML compared to the original DML has been empirically verified in numerical experiments (Singh and Sun, 2019; Chernozhukov et al., 2022a). Thus, extending the procedures and theory of the proposed framework to accommodate signals obtained by Auto-DML is a promising future direction.

The present study proposed the general and flexible framework for the CEFR problems, but more efficient estimation and inference may be possible in some specific settings. For example, the orthogonal moment condition for LATE using the interaction term of Y and D has been proposed in Singh and Sun (2019), while OSR uses Y and D only separately. Comparison of the efficiency of the method in

Singh and Sun (2019) and the proposed one is beyond the scope of this study, but intuitively, leveraging information expressed in the form of interaction of Y and D can improve efficiency. However, the contribution of this study for offering the flexible inference framework in the data combination settings is significant, as the method in Singh and Sun (2019) does not apply to situations where Y and D are separately observed.

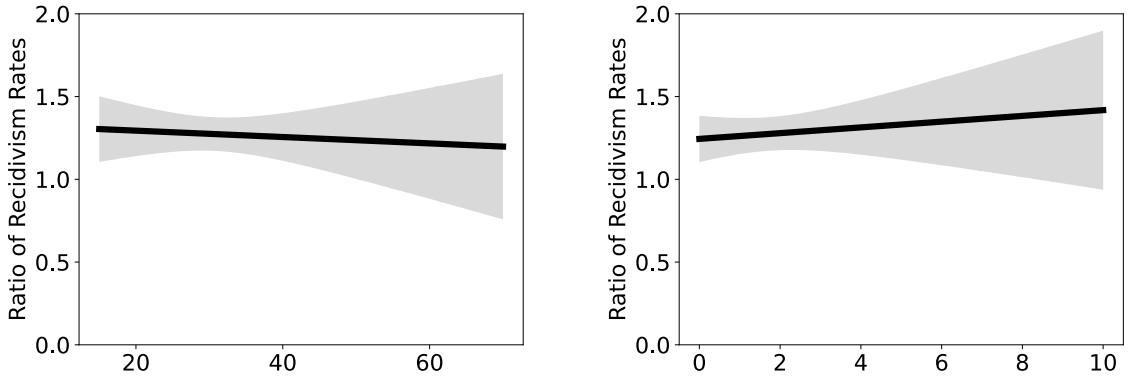
Chapter 4 Estimation of a Treatment Effect from Survival Data with a Cure Fraction Under Insufficient Follow-Up

In survival analysis, there exist many applications where a certain fraction of subjects never experience the event of interest, but the classical survival models suppose that all subjects will experience the event over the course of time. An important and natural example of such applications is the analysis on a relapse risk of a specific disease. Obviously, a proportion of patients will never have the relapse, thus, they are considered cured of their disease. Following such medical applications, the extended survival models that accommodate cured subjects are referred to as *cure models* in the literature. Other examples can be found in a broad range of areas including, but not limited to, sociology (time until recidivism from release), marketing (time until a consumer makes a purchase), and economics (time until a company goes bankrupt and time until the unemployed get a job).

In this study, I consider the estimation of a treatment effect using time-to-event data with a cure fraction. Even though the existence of cured subjects is normally found in various real-world applications, little attention has been paid to causal inference using survival data with a cure fraction. The only previous study I am aware of is by Gao and Zheng (2017). They propose a method for estimating a treatment effect using survival data with a cure fraction and noncompliance. However, they assume RCTs and a parametric model, which raises questions about the usefulness of their method in observational studies. The major contribution of the present study is to propose novel conditions and a method for nonparametric identification and estimation of a treatment effect in observational studies while explicitly assuming a positive cure fraction.

For identification of nonparametric cure models, most previous studies (Laska and Meisner, 1992; Xu and Peng, 2014; López-Cheda et al., 2017a,b; Chown et al., 2020) require the assumption of sufficient follow-up, namely the follow-up time is sufficiently long to observe the event of interest occurs for all uncured subjects. This condition is formally stipulated as

$$\tau_{T_d}(v) < \tau_{C_d}(v) \text{ for } d = 0, 1 \text{ and all } v \in \mathcal{V},$$



(a) Age.

(b) Actual Time Served.

Figure 3: Estimated Ratio of the Recidivism Rates.

where T_d are the potential event time, C_d is the potential censoring time, V is the subvector of the covariate vector X, and $\tau_H(v) = \inf\{t : P(H < t|V = v) = 1\}$. The exceptions are Escobar-Bach and van Keilegom (2019) and Escobar-Bach et al. (2022), which make use of techniques from extreme value theory to nonparametrically estimate a cure fraction under insufficient follow-up, but their methods instead require some auxiliary information that is sometimes difficult to obtain in practice. On the other hand, it is not always necessary to identify the entire cure model when one is interested in a treatment effect, not the cure model itself. I show the ratio of uncured subjects in treatment and control groups:

$$\theta_0(v) := \frac{P(T_1 < \infty|V = v)}{P(T_0 < \infty|V = v)},$$

can be identified even if follow-up is insufficient, and we can use the inference framework proposed in Chapter 3. The proposed method is especially powerful when resources and time for experiments are scarce because it can predict a long-term effect of a treatment without tracing subjects for a long period of time as long as the identification assumptions are satisfied.

A similar idea of using a ratio as a measure of a treatment effect for survival data can be found in Lee (2022), in which the author considers the ratio of survival functions as the measure of a treatment effect in order to cancel out the impact of random censoring. The present study differs from Lee (2022) in that I consider the presence of a cure fraction while Lee (2022) does not, and that I propose to directly estimate the ratio of uncured subjects while Lee (2022) constructs separate estimators for two survival functions.

I apply the proposed method to analyze the effect of the type of release on the recidivism rate using data collected in 11 states in the US. Figure 3(a) and (b) show the estimated ratio of the recidivism rates as a function of the age at release and actual time served, respectively. The grey area in the graphs is

the 95% uniform confidence band. Both graphs indicate that being released on parole or probation is deterrent against recidivism, while the effect is not so strong. This matches the intuition that parole and probation decrease the risk of recidivism because if an ex-inmate gets arrested while on parole or probation, a new prison term is imposed in addition to the rest of the last prison term. However, Bierens and Carvalho (2007) reports that the directions of the effect of parole and probation on the recidivism rate are mixed in the eleven states. It may lead to a small overall effect averaged over the eleven states in this application.

Figure 3(a) implies age weakens relative deterrence of parole and probation against recidivism. Older ex-inmates may be more experienced in living without committing crime or avoiding being caught, which reduces the recidivism rate regardless of the type of release. The uniform confidence band indicates the treatment effect is statistically significantly positive for those younger than 51 years old. Contrary to age, the actual time served strengthens relative deterrence of parole and probation against recidivism, which is also intuitively understandable since we can expect that parolees with a longer sentence are more reluctant to take the risk of rearrest. In Figure 3(b), the treatment effect is statistically significantly positive for those with the actual time served shorter than 8.4 years.

The key identification assumption used in this study is the equal latency:

$$P(T_0 < t | T_0 < \infty, X = x) = P(T_1 < t | T_1 < \infty, X = x),$$

which is relatively strong, and sometimes difficult to verify in practice. In addition to the fact that deep domain knowledge is essential for accurate verification, judgments can easily become subjective. If the equal latency assumption is not satisfied, the proposed method can be biased. Therefore, developing testing methods for the equal latency is necessary for the further advancement of the study on causal inference from survival data with a cure fraction. Furthermore, developing estimation methods based on alternative assumptions other than the equal latency is also an important future direction.

Although the present study only considers covariates that do not change over time, allowing the use of time-varying covariates is highly important when analyzing time-to-event data. However, research on cure models with time-varying covariates is still developing, and there have only been a few previous works (Beretta and Heuchenne, 2019; Dirick et al., 2019; Tawiah et al., 2020; Lambert and Bremhorst, 2020).

References

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

Bareinboim E. and Pearl J. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences of the United States of America*, 113(27):7345–7352, 2016.

Beretta A. and Heuchenne C. Variable selection in proportional hazards cure model with time-varying covariates, application to US bank failures. *Journal of Applied Statistics*, 46(9):1529–1549, 2019.

Bierens H. J. and Carvalho J. R. Semi-nonparametric competing risks analysis of recidivism. *Journal of Applied Econometrics*, 22(5):971–993, 2007.

Brookhart M. A. and van der Laan M. J. A semiparametric model selection criterion with applications to the marginal structural model. *Computational Statistics & Data Analysis*, 50(2):475–498, 2006.

Caron A., Baio G. and Manolopoulou I. Estimating individual treatment effects using non-parametric regression models: A review. *arXiv preprint arXiv:2009.06472*, 2020.

Chernozhukov V., Newey W., Quintas-Martínez V. M. and Syrgkanis V. RieszNet and ForestRiesz: Automatic debiased machine learning with neural nets and random forests. In *Proceedings of the 39th International Conference on Machine Learning*, pages 3901–3914, 2022a.

Chernozhukov V., Newey W. K. and Singh R. Automatic debiased machine learning of causal and structural effects. *Econometrica*, 90(3):967–1027, 2022b.

Chown J., Heuchenne C. and van Keilegom I. The nonparametric location-scale mixture cure model. *TEST*, 29(4): 1008–1028, 2020.

Dirick L., Bellotti T., Claeskens G. and Baesens B. Macro-economic factors in credit risk calculations: Including time-varying covariates in mixture cure models. *Journal of Business & Economic Statistics*, 37(1):40–53, 2019.

Dukes O. and Vansteelandt S. A note on G-estimation of causal risk ratios. *American Journal of Epidemiology*, 187(5):1079–1084, 2018.

Escobar-Bach M. and van Keilegom I. Non-parametric cure rate estimation under insufficient follow-up by using extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 81(5):861–880, 2019.

Escobar-Bach M., Maller R., van Keilegom I. and Zhao M. Estimation of the cure rate for distributions in the Gumbel maximum domain of attraction under insufficient follow-up. *Biometrika*, 109(1):243–256, 2022.

Frölich M. Nonparametric IV estimation of local average treatment effects with covariates. *Journal of Econometrics*, 139(1):35–75, 2007.

Gao X. and Zheng M. Estimating the causal effects in randomized trials for survival data with a cure fraction and non compliance. *Communications in Statistics: Theory and Methods*, 46(8):4065–4087, 2017.

Holland P. W. Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960, 1986.

Imbens G. W. and Angrist J. D. Identification and estimation of local average treatment effects. *Econometrica*, 62 (2):467–475, 1994.

Imbens G. W. and Rubin D. B. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press, 2015.

Lambert P. and Bremhorst V. Inclusion of time-varying covariates in cure survival models with an application in fertility studies. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 183(1):333–354, 2020.

Laska E. M. and Meisner M. J. Nonparametric estimation and testing in a cure model. *Biometrics*, 48(4):1223–1234, 1992.

Lee M. Survival ratio solves random censoring problem for causal effect analysis. *Available at SSRN 4224247*, 2022.

Lee S., Correa J. D. and Bareinboim E. Identifiability from a combination of observations and experiments. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 13677–13680. AAAI Press, Apr. 2020.

Liang M. and Yu M. Relative contrast estimation and inference for treatment recommendation. *arXiv preprint arXiv:2010.13904*, 2020.

López-Cheda A., Cao R., Jácome M. A. and van Keilegom I. Nonparametric incidence estimation and bootstrap bandwidth selection in mixture cure models. *Computational Statistics & Data Analysis*, 105:144–165, 2017a.

López-Cheda A., Jácome M. A. and Cao R. Nonparametric latency estimation for mixture cure models. *TEST*, 26 (2):353–376, 2017b.

Ogburn E. L., Rotnitzky A. and Robins J. M. Doubly robust estimation of the local average treatment effect curve. *Journal of the Royal Statistical Society. Series B (Statistical methodology)*, 77(2):373, 2015.

Ridder G. and Moffitt R. The econometrics of data combination. In Heckman J. J. and Leamer E. E., editors, *Handbook of Econometrics*, volume 6, pages 5469–5547. Elsevier, 2007.

Rolling C. A. and Yang Y. Model selection for estimating treatment effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, pages 749–769, 2014.

Rubin D. B. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688, 1974.

Saito Y. and Yasui S. Counterfactual cross-validation: Stable model selection procedure for causal inference models. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pages 8398–8407. PMLR, 2020.

Schuler A., Baiocchi M., Tibshirani R. and Shah N. A comparison of methods for model selection when estimating individual treatment effects. *arXiv preprint arXiv:1804.05146*, 2018.

Seaman S. R. and White I. R. Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research*, 22(3):278–295, 2013.

Shinoda K. and Hoshino T. Estimation of local average treatment effect by data combination. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 8295–8303, 2022.

Singh R. and Sun L. Automatic debiased machine learning for instrumental variable models of complier treatment effects. *arXiv preprint arXiv:1909.05244*, 2019.

Tawiah R., McLachlan G. J. and Ng S. K. Mixture cure models with time-varying and multilevel frailties for recurrent event data. *Statistical Methods in Medical Research*, 29(5):1368–1385, 2020.

Wooldridge J. M. Inverse probability weighted M-estimators for sample selection, attrition, and stratification. *Portuguese Economic Journal*, 1(2):117–139, 2002.

Wooldridge J. M. Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics*, 141(2):1281–1301, 2007.

Xu J. and Peng Y. Nonparametric cure rate estimation with covariates. *Canadian Journal of Statistics*, 42(1):1–17, 2014.

Yadlowsky S., Pellegrini F., Lionetto F., Braune S. and Tian L. Estimation and validation of ratio-based conditional average treatment effects using observational data. *Journal of the American Statistical Association*, 116(533): 335–352, 2021.

Yamane I., Yger F., Atif J. and Sugiyama M. Uplift modeling from separate labels. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.