*Statistical Science*
2020, Vol. 35, No. 3, 540–544
https://doi.org/10.1214/20-STS796
Main article: https://doi.org/10.1214/20-STS786


# Discussion of "On Nearly Assumption-Free Tests of Nominal Confidence Interval Coverage for Causal Parameters Estimated by Machine Learning"

**Edward H. Kennedy, Sivaraman Balakrishnan and Larry Wasserman**

## 1. INTRODUCTION

We congratulate the authors on their exciting paper, which introduces a novel idea for assessing the estimation bias in causal estimates. Doubly robust estimators are now part of the standard set of tools in causal inference, but a typical analysis stops with an estimate and a confidence interval. The authors give an approach for a unique type of model-checking that allows the user to check whether the bias is sufficiently small with respect to the standard error, which is generally required for confidence intervals to be reliable.

We begin our comments by looking at an example of a simple functional.

## 2. EXPECTED DENSITY EXAMPLE

In this section, we illustrate the main ideas in the paper by applying them to a simpler functional. This allows us to understand better some of the critical insights of Liu, Mukherjee and Robins (2020). In particular, we consider the classic expected density functional

$$\psi = \mathbb{E}\{p(X)\} = \int p(x)^2\,dx.$$

This functional has been studied extensively, with estimation and inference by now well understood (Bickel and Ritov, 1988, Birgé and Massart, 1995). Further, although it is simple, it has many of the nice properties of more complicated functionals like the expected conditional covariance or average treatment effect.

An analog of a doubly robust estimator of $\psi$ is the one-step or first-order corrected estimator given by

$$\widehat{\psi} = \frac{2}{n}\sum_{i=1}^{n} \widehat{p}(X_i) - \int \widehat{p}(x)^2\,dx,$$

where $\widehat{p}$ is an initial pilot estimator of the density $p$, which for simplicity is based on an independent auxiliary sample of size $n$. The rest of this analysis is conditioned on this auxiliary sample.

To fix ideas, we briefly summarize some results regarding the estimation of $\psi$ and the estimator $\widehat{\psi}$. A simple calculation shows that we may write

$$\psi = \widehat{\psi} - 2\left(\frac{1}{n}\sum_{i=1}^{n} \widehat{p}(X_i) - \mathbb{E}[\widehat{p}]\right) + \int (\widehat{p} - p)^2.$$

In rough terms, if $\int(\widehat{p}-p)^2$ is $o_p(1/\sqrt{n})$ then the first-order estimator achieves parametric rates (and is semiparametrically efficient). As an example, over classical Sobolev or Hölder smoothness classes, the first-order estimator is efficient if $s > d/2$, where $s$ denotes the smoothness parameter, and $d$ the dimension of the data. On the other hand, it is well known that a second-order U-statistic estimator (Laurent, 1996) is semiparametrically efficient if $s > d/4$ and otherwise achieves the minimax rate of $n^{-4s/(4s+d)}$.

To understand the work of Liu, Mukherjee and Robins (2020), suppose we write our initial estimate as

$$\widehat{p} = \sum_{j=1}^{\infty} \widehat{\theta}_j \phi_j,$$

where the $\phi_j$ form an orthonormal basis with respect to the Lebesgue measure. Then for $p = \sum_j \theta_j \phi_j$ a straightforward calculation shows that the conditional bias (given the auxiliary sample) is

$$\text{Bias} = \mathbb{E}(\widehat{\psi} - \psi) = -\int \{\widehat{p}(x) - p(x)\}^2\,dx$$
$$= -\sum_{i=1}^{\infty} (\widehat{\theta}_j - \theta_j)^2.$$

Note that this is the bias, not the squared bias; therefore for this functional, the standard first-order estimator has the monotone bias property. We can decompose this bias as

$$-\text{Bias} = \underbrace{\sum_{j=1}^{k} (\widehat{\theta}_j - \theta_j)^2}_{\text{Bias}_k} + \underbrace{\sum_{j>k} (\widehat{\theta}_j - \theta_j)^2}_{\text{Truncation Bias}}.$$

*Edward H. Kennedy is Assistant Professor, Department of Statistics & Data Science, Carnegie Mellon University, USA (e-mail: edward@stat.cmu.edu). Sivaraman Balakrishnan is Assistant Professor, Department of Statistics & Data Science, Carnegie Mellon University, USA (e-mail: siva@stat.cmu.edu). Larry Wasserman is University Professor, Department of Statistics & Data Science, Carnegie Mellon University, USA (e-mail: larry@stat.cmu.edu).*