

1 Summary

I propose an extension to the Double Machine Learning (DML) estimator, originally developed in Chernozhukov et al. (2018), to the case of multiple outcomes of interest, where correlations between equations are taken into account. The model framework will be a non-parametric version of the seemingly unrelated regression (SUR) model, and the newly proposed estimator will be preliminarily called the General Double Machine Learning (GDML) estimator. Advantages of studying multiple outcomes simultaneously include potential drastic improvements in efficiency over separate models, and asymptotically valid joint hypothesis testing of multiple causal objects of interest. Necessary asymptotic and finite sample results will be derived to establish the properties of the new estimator, as well as to construct an estimator of the covariance matrix of the estimator. The validity and advantages of the GDML estimator will be evaluated with Monte Carlo simulations and an empirical application.

2 Introduction

Recent developments have shown that machine learning methods can be used effectively for a variety of causal inference problems. In particular, Chernozhukov et al. (2018) developed the Double/Debiased machine learning (DML) estimator, which is capable of estimating a variety of causal objects of interest with the proper construction of moment conditions. For example, the DML estimator allows for the estimation of the slope coefficient in the partial linear model, and also the average treatment effect (ATE) for the binary case (the ATE estimation has since been extended to the continuous case in Su et al. (2019) and Colangelo and Lee (2020)).

Consider an outcome of interest Y , treatment T (allowed to be multi-dimensional), co-

variate set X , and error ε . We may be interested in a nonparametric outcome equation

$$Y = g(T, X) + \varepsilon, \quad (1)$$

which also encompasses parametric equations as special cases (such as the partial linear model from Robinson (1988) considered in Chernozhukov et al. (2018)). The DML estimator requires a first-stage estimation of the conditional mean function $\gamma(t, x) = \mathbb{E}(Y|T = t, X = x)$ and the propensity score $f_{T|X}(t|x)$, typically using machine learning methods.¹

In many applications we may wish to study the relationship between treatments and multiple different outcomes of interest. If we consider m outcomes of interest, we can extend equation 1 for all m outcomes:

$$\begin{aligned} Y_1 &= g_1(T, X) + \varepsilon_1, \\ Y_2 &= g_2(T, X) + \varepsilon_2, \\ &\vdots \\ Y_m &= g_m(T, X) + \varepsilon_m, \end{aligned}$$

where each g_j is allowed to be a different function which may be dependent on a different subset of covariates. The m equations can be stacked together as

$$\mathbf{Y} = \mathbf{g}(T, X) + \varepsilon, \quad (2)$$

where $\mathbf{Y} = (Y_1, \dots, Y_m)'$ and \mathbf{g} is a vector valued function. In the strictly linear case this would be equivalent to the Seemingly Unrelated Regression (SUR) model as in Zellner (1962). I am unaware of any research on nonparametric SUR models as specified in equation 2 from a frequentist perspective (Smith and Kohn (2000) propose a Bayesian estimator but it relies

¹Note that any method that satisfies the convergence assumptions are allowed, including nonparametric kernel estimators or even linear regression

on linearity with respect to a set of basis function). While the SUR model can be estimated equation by equation, it has been shown in prior literature such as Zellner (1963) that estimating an SUR model jointly via Generalized Least Squares (GLS) can result in efficiency gains. Further efficiency gains have been demonstrated for SUR models in Mehrabani and Ullah (2020) with the use of restricted estimation. Additionally, analyzing the outcomes jointly allows for testing joint hypotheses on parameters relating to different outcomes. This indicates that there may be some advantages to applying analogous concepts to the DML estimator.

In the case of equation 2, an implementation of the DML estimator would necessitate the estimation of nuisance parameters γ_j and $f_{T|X}$ independently for $j = 1, \dots, m$, which would then be used to estimate the moment condition(s) for each equation independently. However this ignores any correlation between equations, which it would be reasonable to expect (for example we would expect covariates most relevant in one model to also be very relevant in another model). There is a growing literature on multi-output (or multi-label) machine learning algorithms that take advantage of such correlations (see Borchani et al. (2015) for a survey of such algorithms). There is also a fast growing literature on transfer learning and prior knowledge learning, in which information learned about one equation is directly inputted into another (for example the Discriminability-Based Transfer (DBT) algorithm in Pratt (1993) or the prior lasso described in Jiang et al. (2016)). Usage of such algorithms would allow for the joint estimation of each γ_j which can improve performance of the nuisance estimator.

To take advantage of the potential correlation between equations I propose the General Double Machine Learning (GDML) estimator which consists of two components: The joint estimation of the first-stage nuisance parameters (γ_j and $f_{T|X}$) for all m equations and an averaging estimator similar to Mehrabani and Ullah (2020). As a special case the GDML estimator will encompass equation-by-equation DML.

3 Goals For the Summer

Over the summer I hope to accomplish the following goals:

1. Establish necessary theoretical results for the proposed estimator.
2. Code and run Monte Carlo simulations to assess the performance of the proposed estimator.
3. Develop a novel empirical application of the new estimator.
4. Complete a draft of the paper and proceed with the dissertation prospectus exam.
5. Submit the paper to a number of conferences

4 Budget

I hereby request the following budget for my project:

- \approx \$4,000 for Amazon Web Services for the intense computation required for the Monte Carlo simulations.

Any funding I do receive will necessarily go towards these computational costs.

References

- Borchani, Hanen, Gherardo Varando, Concha Bielza, and Pedro Larrañaga**, “A survey on multi-output regression,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2015, 5 (5), 216–233.
- Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, and Whitney Newey**, “Double/debiased/neyman machine learning of treatment effects,” *American Economic Review*, 2017, 107 (5), 261–65.
- , – , – , – , – , – , – , **and James Robins**, “Double/debiased machine learning for treatment and structural parameters,” 2018.
- Colangelo, Kyle and Ying-Ying Lee**, “Double debiased machine learning nonparametric inference with continuous treatments,” *arXiv preprint arXiv:2004.03036*, 2020.
- Jiang, Yuan, Yunxiao He, and Heping Zhang**, “Variable selection with prior information for generalized linear models via the prior LASSO method,” *Journal of the American Statistical Association*, 2016, 111 (513), 355–376.
- Mehrabani, Ali and Aman Ullah**, “Improved Average Estimation in Seemingly Unrelated Regressions,” *Econometrics*, 2020, 8 (2), 15.
- Pratt, Lorien Y**, “Discriminability-based transfer between neural networks,” in “Advances in neural information processing systems” 1993, pp. 204–211.
- Robinson, Peter M**, “Root-N-consistent semiparametric regression,” *Econometrica: Journal of the Econometric Society*, 1988, pp. 931–954.
- Smith, Michael and Robert Kohn**, “Nonparametric seemingly unrelated regression,” *Journal of Econometrics*, 2000, 98 (2), 257–281.
- Su, Liangjun, Takuya Ura, and Yichong Zhang**, “Non-separable models with high-dimensional data,” *Journal of Econometrics*, 2019, 212 (2), 646–677.

Zellner, Arnold, “An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias,” *Journal of the American statistical Association*, 1962, 57 (298), 348–368.

—, “Estimators for seemingly unrelated regression equations: Some exact finite sample results,” *Journal of the American Statistical Association*, 1963, 58 (304), 977–992.