Debiased Machine Learning (DML)

Alexander Quispe

November 20, 2021

Citation

These notes are based on the Lecture Notes of Victor Chernozukhov ML and Causal Inference Course - 2021.

DML Inference in Partially Linear Regression Model (PLM)

We first answer the questions posed above within the context of the partially linear regression model, which reads:

$$Y = \beta D + g(X) + \epsilon, \quad \mathrm{E}[\epsilon \mid X, D] = 0, \tag{2.1}$$

where Y is the outcome variable, D is the regressor of interest, and X is a high-dimensional vector of other regressors or features, called "controls". The coefficient β provides the answer to the predictive effect question. In this segment we discuss estimation and confidence intervals for β. We also provide a case study, in which we examine the effect of gun ownership on homicide rates.

DML Inference in Partially Linear Regression Model (PLM)

In what follows, we will employ the partialling out X operation of the form that inputs a random variable V and outputs the residualized form:

$$\tilde{V} := V - \mathrm{E}[Y \mid X].$$

Applying this operation to (2.1) we obtain:

$$\tilde{Y} = \beta\tilde{D} + \epsilon, \quad \mathrm{E}(\epsilon\tilde{D}) = 0, \tag{2.2}$$

where $\tilde{Y}$ and $\tilde{D}$ are the residuals left after predicting Y and D using X. Spefically, we have that

$$\tilde{Y} := Y - \ell(X), \quad \tilde{D} := D - m(X),$$

DML Inference in Partially Linear Regression Model (PLM)

Theorem 2.1 (FWL Partialling-Out for Partially Linear Model) *Suppose that Y, X and D have bounded second moments. Then population regression coefficient β can be recovered from the population linear regression of $\tilde{Y}$ on $\tilde{D}$:*

$$\beta := \{b : \mathrm{E}(\tilde{Y} - b\tilde{D})\tilde{D} = 0\} := (\mathrm{E}\tilde{D}^2)^{-1}\mathrm{E}\tilde{D}\tilde{Y},$$

DML Inference in Partially Linear Regression Model (PLM)

Double/Orthogonal ML for Partially Linear Model

1. Partition data indices into random folds of approximately equal size: $\{1, \ldots, n\} = \cup_{k=1}^{K} I_k$. For each fold $k = 1, \ldots, K$, compute ML estimators $\hat{\ell}_{[k]}$ and $\hat{m}_{[k]}$ of the conditional expectation functions ℓ and m, leaving out the k-th block of data, and obtain the cross-fitted residuals or each $i \in I_k$:

$$\check{Y}_i = Y_i - \hat{\ell}_{[k]}(X_i), \quad \check{D}_i = D_i - \hat{m}_{[k]}(X_i).$$

2. Apply ordinary least squares of $\check{Y}_i$ on $\check{D}_i$, that is, obtain the $\hat{\beta}$ as the root in b of the normal equations:

$$\mathbb{E}_n(\check{Y}_i - b\check{D}_i)\check{D}_i = 0.$$

3. Construct standard errors and confidence intervals as in the standard least squares theory.

DML Inference in Partially Linear Regression Model (PLM)

Selecting the Best ML Learners of ℓ and m. There may be several methods that satisfy the quality requirements of the theorem stated above, and we may therefore ask the question of what ML methods we should use in practice. Consider a collection of ML methods indexed by $j \in \{1, \ldots, J\}$, our goal would be to select the methods that minimize an upper bound on the bias of the DML estimator.

DML Inference in Partially Linear Regression Model (PLM)

Selecting the Best ML Learners of ℓ and m. There may be several methods that satisfy the quality requirements of the theorem stated above, and we may therefore ask the question of what ML methods we should use in practice. Consider a collection of ML methods indexed by $j \in \{1, \ldots, J\}$, our goal would be to select the methods that minimize an upper bound on the bias of the DML estimator.

DML Inference in Partially Linear Regression Model (PLM)

the mean square approximation errors (MSAE):

$$\frac{1}{K}\sum_{k=1}^{K} ||\hat{\ell}_{[k]} - \ell||^2_{L^2} \text{ and } \frac{1}{K}\sum_{k=1}^{K} ||\hat{m}_{[k]} - m||^2_{L^2}. \qquad (2.3)$$

DML Inference in Partially Linear Regression Model (PLM)

Selection of the Best ML Methods for DML to Minimize Bias. Consider a set of ML methods enumerated by $j \in \{1, \ldots, J\}$.

- For each method j, compute the cross-fitted MSPEs

$$\mathbb{E}_n \check{Y}_{i,j}^2 \text{ and } \mathbb{E}_n \check{D}_{i,j}^2,$$

  where the index j reflects the dependency of residuals on the method.

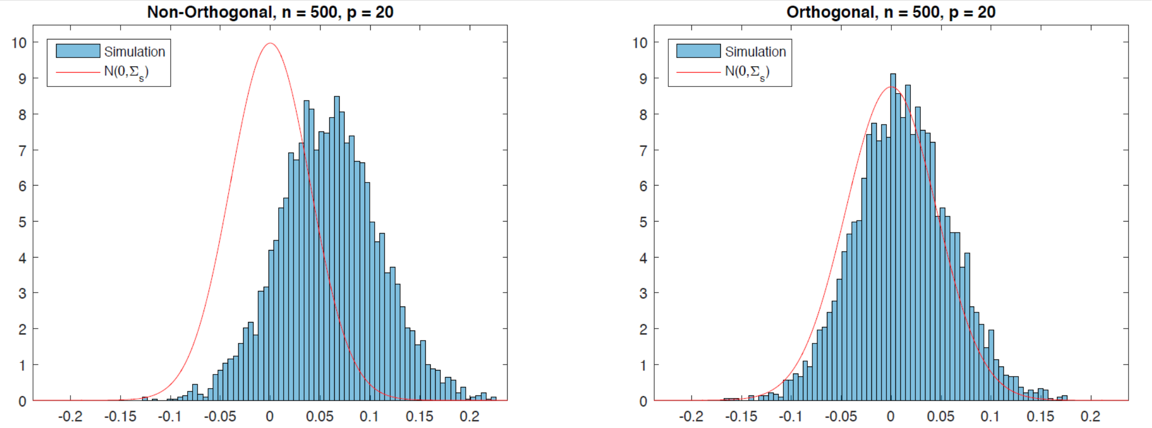
- Select the ML methods $j \in \{1, \ldots, J\}$ that give the smallest MSPEs:

$$\hat{j}_\ell = \arg\min_j \mathbb{E}_n \check{Y}_{i,j}^2 \text{ and } \hat{j}_m = \arg\min_j \mathbb{E}_n \check{D}_{i,j}^2.$$

- Use the method $\hat{j}_\ell$ as a learner of ℓ, and $\hat{j}_m$ as a learner of m in the DML algorithm above.

DML Inference in Partially Linear Regression Model (PLM)

Figure: 1

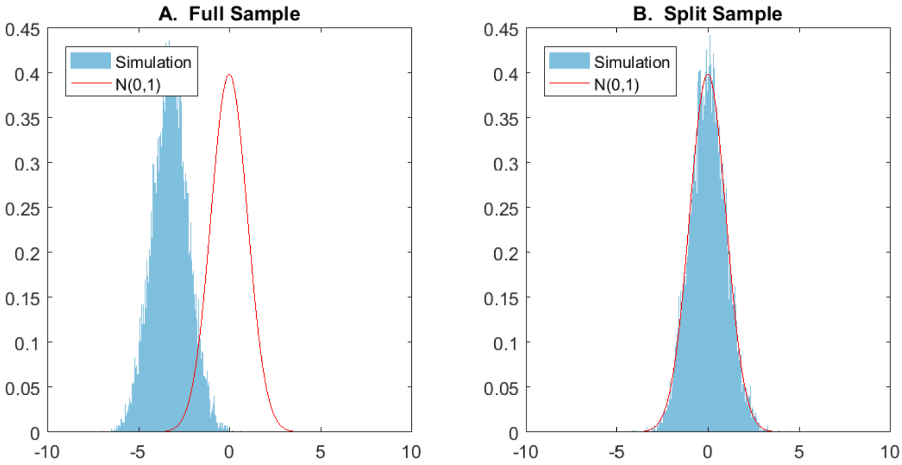


DML Inference in Partially Linear Regression Model (PLM)

Figure 1, left panel, illustrates the bias arising due to the use of non-orthoghonal, naive learning of β. Specifically,the figure shows the behavior of a conventional (non-orthogonal) ML estimator, $\tilde{\beta}$, in the partially linear model in a simple simulation experiment where we learn g using a random forest. The g in this experiment is a very smooth function of a small number of variables, so the experiment is seemingly favorable to the use of random forests a priori. The histogram

DML Inference in Partially Linear Regression Model (PLM)

Figure: 2



DML Inference in Partially Linear Regression Model (PLM)

Figure 2 illustrates how the bias resulting from overfitting in the estimation of nuisance functions can cause the DML (without sample splitting) to be biased and how sample splitting completely eliminates this problem. In the left panel the histogram shows the finite-sample distribution of the DML in the partially linear model where nuisance parameters are estimated with overfitting using the full sample, i.e. without sample splitting. The finite-sample distribution is clearly

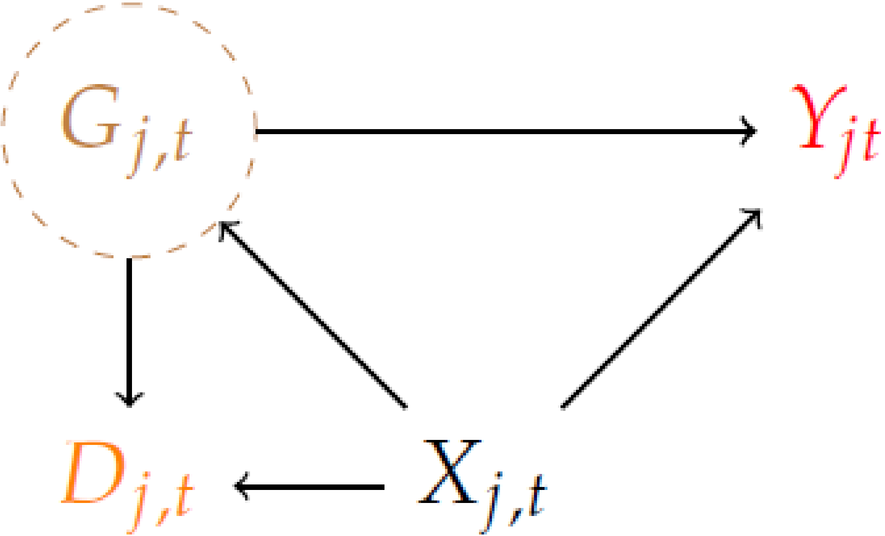
The Effect of Gun Ownership on Gun-Homicide Rates

We consider the problem of estimating the effect of gun ownership on the homicide rate. For this purpose, we estimate the partially linear model:

$$Y_{j,t} = \beta D_{j,(t-1)} + g(X_{j,t}) + \epsilon_{j,t}.$$

$Y_{j,t}$ is log homicide rate in county j at time t, $D_{j,t-1}$ is log fraction of suicides committed with a firearm in county j at time $t-1$, which we use as a proxy for gun ownership $G_{j,t}$, which is not observed, and $X_{j,t}$ is a set of demographic and economic characteristics of county j at time t.

The Effect of Gun Ownership on Gun-Homicide Rates



The Effect of Gun Ownership on Gun-Homicide Rates

	Estimate	Standard Error
Baseline OLS	0.282	0.065
Least Squares with controls	0.191	0.052
Lasso	0.223	0.057
Post-Lasso	0.227	0.056
CV Lasso	0.200	0.058
CV Elnet	0.206	0.057
CV Ridge	0.201	0.058
Random Forest	0.192	0.058
DNN	0.176	0.116
Best	0.217	0.058

The Effect of Gun Ownership on Gun-Homicide Rates

The last row of the table provides the "best" estimates. To obtain "best" estimates we evaluate the performance of predictors $\hat{\ell}(X)$ and $\hat{m}(X)$ estimated by different methods on

The Effect of Gun Ownership on Gun-Homicide Rates

The last row of the table provides the "best" estimates. To obtain "best" estimates we evaluate the performance of predictors $\hat{\ell}(X)$ and $\hat{m}(X)$ estimated by different methods on

The Effect of Gun Ownership on Gun-Homicide Rates

auxiliary samples using the main sample. Then we pick the methods giving the lowest MSE. In our case ridge regression and lasso give the best performances in predicting $Y_{j,t}$ and $D_{j,t-1}$, respectively. We then use the best methods as predictors in estimation procedure described above. The resulting estimate of the gun ownership effect and standard error are similar to that of Lasso.