Answers to Verbal Questions - Empirical Project I

ECON 2390

October 14, 2024

Question 5

The randomization assumption is largely reasonable based on the results, as most covariates appear well-balanced between the treated and control groups. However, there is evidence of an imbalance in the 'degree' and 'education' covariates, which may warrant a further adjustment in the analysis to account for these differences.

Question 7

Using Random Forest, we relax the parametric assumption about the relationship between the covariates and the treatment assignment. The Random Forest model can automatically capture non-linear relationships and complex interactions among covariates. The main cost in Random Forest models is that we can't easily interpret how individual covariates affect the treatment assignment.

Question 8

Score distributions are similar across both methods, as the density curves largely overlap. However, the overlap between the treated and control groups is limited, as the propensity scores concentrated near 0 for the control group.

Question 9

The Probit and Random Forest models perform well, with substantial treatment effects observed (SMD > 1) and highly significant p-values across ATT and ATE estimators. The similarity in results is reassuring, suggesting that both methods and estimators effectively capture the treatment effects. While the Random Forest model offers greater flexibility in capturing non-linear relationships and interactions, the Probit model remains valuable for its simplicity and interoperability.

Question 10

Yes, the ATE and ATT estimators are unbiased under the assumption of unconfoundedness. Typically, when ATE > ATT, it suggests that the treatment effect is larger for the broader population than for the treated group alone. In this case, however, the small difference between ATE and ATT indicates that the treatment effect is consistent across both the treated group and the broader population, implying a relatively homogeneous treatment effect.

Question 11

Unlike the IPW estimator, the DR ensures that the estimator remains unbiased if the propensity score or outcome models are correctly specified. This flexibility allows for using nonparametric methods like Random Forest, as performed here. To minimize bias, the machine learning models must avoid overfitting, ensuring they generalize well to new data. Additionally, for the estimator to achieve $\sqrt{N}$-asymptotic normality, the models must satisfy the product rate condition, which requires both the propensity score and outcome models to converge at appropriate rates.

Question 12

By partitioning the data into several folds, cross-fitting reduces the statistical dependence between the nuisance functions (such as propensity scores and outcome models) and the ATE estimator. The estimator achieves $\sqrt{N}$-asymptotic normality if the machine learning models satisfy the product rate condition.

Question 13

The Doubly Robust (DR) estimator is unbiased if the propensity score or outcome models are correctly specified. The cross-fitted DR estimator inherits this property while reducing the statistical dependence between the nuisance function, thereby achieving $\sqrt{N}$-asymptotic normality. Hence, the Cross-Fitted DR Estimator is the preferred estimator.