Tests for Heterogeneous Treatment Effect

PhD conference

Fangzhou Yu

2023-11-16

Overview

- Propose semiparametric tests for the **presence of Heterogeneous Treatment Effect (HTE)**
- Two-step procedure:
  1. **Augmented Inverse Propensity Weighting (AIPW)** transformation on the outcome
  2. Project the transformed outcome on the covariates, test **joint significance of projection coefs**
- Test conditional ATE/LATE in the same framework

$$\mathbb{E}[Y_i(1) - Y_i(0)|X_i = x], \qquad \mathbb{E}[Y_i(1) - Y_i(0)|\underbrace{W_i(Z_i = 1) = 1}_{\text{compliers}}, X_i = x]$$

- Combine the tests with Double/Debiased Machine Learning (DML) to deal with high-dimensional covariates

Setting and Hypotheses for CATE

Potential Outcome Framework

- Treatment status: $W_i = 0, 1$
- Covariates: X_i
- Potential outcomes for individual i: $Y_i(1)$ if treated, $Y_i(0)$ if untreated
- Observed outcome: $Y_i = W_i Y_i(1) + (1 - W_i) Y_i(0)$
- CATE: $\tau(x) = \mathbb{E}[Y_i(1) - Y_i(0) | X_i = x]$
- ATE: $\bar{\tau} = \mathbb{E}[\tau(X_i)]$

Two pairs of hypotheses:

1. $H_0 : \tau(x) = 0$ for all x **vs.** $H_a : \tau(x) \neq 0$ for some x
2. $H_0 : \tau(x)$ is constant for all x **vs.** $H_a : \tau(x)$ is not constant for some x

Identification of CATE

Identification Assumptions for CATE

Unconfoundedness: $W_i \perp (Y_i(1), Y_i(0)) | X_i$

Overlap: $\exists \xi > 0$, s.t. $\xi \leqslant e_0(x) \leqslant 1 - \xi$ for all x

where $e_0(x) = P(W_i = 1 | X_i = x)$ is the propensity score.

- Further define $\mu_0(w, x) = \mathbb{E}[Y_i(w) | X_i = x]$
- Then the AIPW transformation is

$$\hat{Y}_i^* = W_i \frac{Y_i - \hat{\mu}(1, X_i)}{\hat{e}(X_i)} - (1 - W_i) \frac{Y_i - \hat{\mu}(0, X_i)}{1 - \hat{e}(X_i)} + \hat{\mu}(1, X_i) - \hat{\mu}(0, X_i)$$

- Under identification assumptions, $\mathbb{E}[\hat{Y}_i^* | X_i = x] = \tau(x)$

Related Literature

- This paper studies the **same hypotheses as in Crump et al. (2008)**
  - But use different methods and extend to heterogeneity in LATE
- Employ the idea of inference on **conditional effects with IPW**
  - Abrevaya, Hsu, and Lieli (2015); Chang, Lee, and Whang (2015); Hsu (2017); **Sant'Anna (2021)**

The closest paper to mine is **Sant'Anna (2021)**:

- **Sant'Anna (2021)**: distributional besides average effects, and allow for right-censored data, which I do not have
- On the other hand, I have the unique features:
  - Tests are **straightforward to implement** by using AIPW instead of IPW, as only standard outputs from statistical softwares are required
  - Develop tests with DML when **identification is achieved with high-dimensional covariates**
  - **Projection coefficients** of conditional effects are estimated, which can provide a profile of the heterogeneity

Why AIPW?

- In the literature, IPW is commonly used to identify CATE
- What is the benefit of using AIPW?

1. Semiparametric Efficiency
    - Related to test power
2. Simplify the formula of asymptotic variance
    - Easier to implement in practice
3. **The resulting moment is Neyman Orthogonal, which is a key condition for using DML**
    - Extension to high-dimensional controls

AIPW: Parametric Estimation

Two-step procedure

1. Transform Y_i by **AIPW** with consistent **parametric** estimators $\hat{\mu}(w, x)$ and $\hat{e}(x)$ to obtain $\hat{Y}_i^*$
2. Estimate the projection coefficients $\beta_0 = (\beta_0^c, \beta_0^{x'})'$ of $\hat{Y}_i^*$ on X_i by OLS.

$$\hat{\beta} = \arg\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} (\hat{Y}_i^* - X_i \beta)^2$$

Then the two pairs of hypotheses translate into

$$\begin{aligned} H_0 : \tau(x) = 0 \quad & \rightarrow \quad H_0 : \beta_0 = 0 \\ H_0 : \tau(x) \text{ is constant} \quad & \rightarrow \quad H_0 : \beta_0^x = 0 \end{aligned}$$

AIPW: Inference

- The key to the hypothesis tests is deriving the asymptotic distribution of $\hat{\beta}$
- Need to take into account plugged-in $\hat{e}(x)$ and $\hat{\mu}(w, x)$

$$\sqrt{n}(\hat{\beta} - \beta_0) = \left(\frac{1}{n} \sum_{i=1}^{n} X_i' X_i \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(X_i \epsilon_i + \underbrace{X_i' \frac{\partial \hat{Y}_i^*}{\partial \gamma} (\hat{\gamma} - \gamma_0)}_{\frac{1}{\sqrt{n}} \sum \cdot \to 0 \text{ because of AIPW}} \right) \right)$$

where γ are the nuisance parameters in $e(x)$ and $\mu(w, x)$. Thus, $\hat{\beta}$ has the **standard OLS asymptotic distribution**

Proposition 3.2(Asymptotic Distribution)

Under regularity conditions and $\hat{\beta} \xrightarrow{p} \beta_0$, $\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V)$ where

$$V = \mathbb{E}[X_i' X_i]^{-1} \mathbb{E}[X_i' X_i \epsilon_i^2] \mathbb{E}[X_i' X_i]^{-1}$$

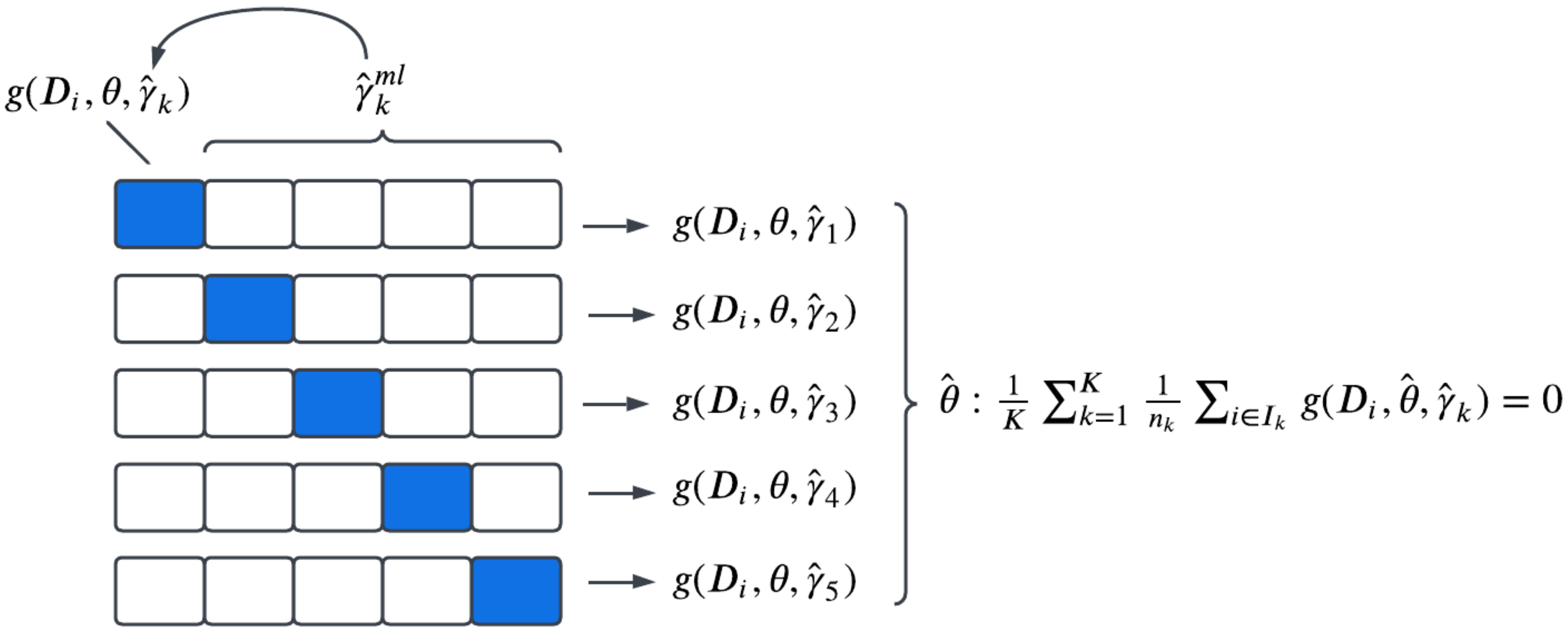
Extension to DML

DML procedure:

1. Transform Y_i by AIPW with **ML estimators** $\hat{e}(x)$ and $\hat{\mu}(w,x)$. A wide range of ML methods can be applied.

2. The OLS estimators $\hat{\beta}$ with **cross-fitting** has the standard asymptotic distribution as in the parametric case.

Cross-fitting:



Extension to DML

Two key conditions to use DML:

1. **Neyman Orthogonality**

   I prove that the moment function of projecting **AIPW** transformed $\hat{Y}_i^*$ on X_i is **Neyman Orthogonal**, namely

$$\partial_\gamma \mathbb{E}[g(D_i, \beta, \gamma_0)][\gamma - \gamma_0] = 0$$

   where ∂_γ is the Gatueax derivative operator and $\gamma = (e(x), \mu(w, x))$ is infinite-dimensional parameter with

$$g(D_i, \beta, \gamma) = X_i'(\hat{Y}_i^* - X_i\beta)$$

2. Provide regularity conditions on the **moment** and on **ML estimator** $\hat{\gamma}$

Wald Test of Heterogeneity

Thus, given the asymptotic distribution and a consistent estimator of the variance, the tests are implemented by **Wald tests of joint significance**

Theorem 3.1(Hypothesis Tests for CATE)

Assume Unconfoundedness and Overlap holds, under $H_0 : (\beta_0^c, \beta_0^{x'})' = 0$, the Wald statistic

$$W_1 = (\hat{\beta}_c, \hat{\beta}_x')(\hat{V}/n)^{-1}(\hat{\beta}_c, \hat{\beta}_x') \xrightarrow{d} \chi^2(p+1)$$

and under $H_0 : \beta_0^x = 0$, the Wald statistic

$$W_2 = \hat{\beta}_x'(\hat{V}_{xx}/n)^{-1}\hat{\beta}_x \xrightarrow{d} \chi^2(p)$$

Setting and Hypotheses for CLATE

- When Unconfoundedness does not hold and a binary IV is available, conditional LATE is identified by

$$CLATE(x) = \frac{\mathbb{E}[\hat{Y}_i^* | X_i = x]}{\mathbb{E}[\hat{W}_i^* | X_i = x]}$$

- Z_i is the "treatment" in the AIPW transformation of both Y_i and W_i
- Let β_0 and α_0 denote the BLP coefficients in the numerator and denominator respectively
- The analogous hypotheses for CLATE:
  - $H_0 : CLATE(x) = 0$ for all x translates into $\beta_0 = 0$
  - H_0 : constant $CLATE(x)$ translates into $\beta_0^x = \frac{\beta_0^c}{\alpha_0^c} \alpha_0^x$

Obtain $\hat{\alpha}$ and $\hat{\beta}$ by one regression, and implement **Wald test + Delta method**

$$\begin{bmatrix} \hat{Y}^* \\ \hat{W}^* \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \mathbf{0} & X & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & X \end{bmatrix} \begin{bmatrix} \beta_c \\ \alpha_c \\ \beta_x \\ \alpha_x \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}$$

Simulated Size of the Tests

Designs | Result

$$Y_i(w) = \mu_0(0, X_i) + \tau_0(X_i) \cdot w + \epsilon_i, \quad Pr(W_i = 1|X_i) = e_0(X_i)$$

Designs:

1. $\mu_0(0, x) = 1 + \sum_{p=1}^{2} x_p; \tau_0(x) = \alpha \sum_{p=1}^{2} x_p; e_0(x) = 0.5$
2. $\mu_0(0, x) = 1 + \sum_{p=1}^{2} x_p; \tau_0(x) = \alpha \sum_{p=1}^{2} x_p; e_0(x) = 1/(1 + \exp(-1 - 0.4x_1 + 0.2x_2))$
3. $\mu_0(0, x) = 1 + \sum_{p=1}^{2} x_p; \tau_0(x) = \alpha(x_1 + x_2 + x_1 \cdot x_2); e_0(x) = 1/(1 + \exp(-1 - 0.4x_1 + 0.2x_2))$
4. $\mu_0(0, x) = 1 + \sum_{p=1}^{2} x_p; \tau_0(x) = \alpha(1\{x_1 > 0\} - 1\{x_1 \leqslant 0\}); e_0(x) = 1/(1 + \exp(-1 - 0.4x_1 + 0.2x_2))$

IV: Linear outcome and TE functions similar to Design 2, but with a binary instrument that has a logit propensity score

Application

CATE Tests   BLP coefficients

Effect of being the only children on mental health in China

- Data constructed by Xie and Hu (2014), analyzed by Zeng, Li, and Ding (2020)
- W_i: Being the only child
- Y_i: Confidence, Anxiety, Desperation ranging 1 (worst) to 5 (best)
- Covariates: Personal and parents' socio-demographic features

- Zero CATE and Constant CATE **rejected** for all three mental measures at 1% level
- ATE estimates: $-0.088^{***}, 0.005, -0.067^{***}$
  - Being the only children decreases confidence and exacerbates desperation
  - The insignificance of the effect on anxiety can be **misleading if focused on ATE only**

Final Notes

- Rejecting H_0 is sufficient but **not** necessary for the existence of HTE
  - The tests are based on linear projection
- **Recommendation:** Add higher-order terms to test for potential non-linear HTE

- Works with high-dimensional identification but **not** high-dim heterogeneity
  - With high-dim covariates, need to project TE on a low-dim subset
  - Regularity conditions are binding the choice of this low-dim subset
- **Recommendation:** Constant CATE test with respect to each covariate, one by one
  - Convert the high-dim HTE to a multiple testing problem
  - List, Shaikh, and Xu (2019): a bootstrap-based approach to the multiple testing problem allowing for dependence among test statistics

Conclusion

- Two-step tests for HTE based on linear projection of HTE on covariates
- Straightforward to implement and only require built-in functions in software/packages
- Flexible and unified method for CATE/CLATE/high-dim cases
- The applications illustrate the use of the tests and its assistance to subpopulation analysis

References

Abrevaya, Jason, Yu-Chin Hsu, and Robert P Lieli. 2015. "Estimating Conditional Average Treatment Effects." *Journal of Business & Economic Statistics* 33 (4): 485–505.

Chang, Minsu, Sokbae Lee, and Yoon-Jae Whang. 2015. "Nonparametric Tests of Conditional Treatment Effects with an Application to Single-Sex Schooling on Academic Achievements." *The Econometrics Journal* 18 (3): 307–46.

Crump, Richard K, V Joseph Hotz, Guido W Imbens, and Oscar A Mitnik. 2008. "Nonparametric Tests for Treatment Effect Heterogeneity." *The Review of Economics and Statistics* 90 (3): 389–405.

Hsu, Yu-Chin. 2017. "Consistent Tests for Conditional Treatment Effects." *The Econometrics Journal* 20 (1): 1–22.

List, John A, Azeem M Shaikh, and Yang Xu. 2019. "Multiple Hypothesis Testing in Experimental Economics." *Experimental Economics* 22: 773–93.

Sant'Anna, Pedro HC. 2021. "Nonparametric Tests for Treatment Effect Heterogeneity with Duration Outcomes." *Journal of Business & Economic Statistics* 39 (3): 816–32.

Xie, Yu, and Jingwei Hu. 2014. "An Introduction to the China Family Panel Studies (CFPS)." *Chinese Sociological Review* 47 (1): 3–29.

Zeng, Shuxi, Fan Li, and Peng Ding. 2020. "Is Being an Only Child Harmful to Psychological Health?: Evidence from an Instrumental Variable Analysis of China's One-Child Policy." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 183 (4): 1615–35.