Testing the LATE Consequence that Only Compliers Move*

Gaurav Sood

September 10, 2025

Abstract

Randomized encouragement designs often feature imperfect compliance. Local average treatment effects (LATE) identify causal effects under random assignment, exclusion, and monotonicity. These conditions imply a strong distributional consequence: when the instrument flips, only compliers can move the outcome distribution. This paper turns that consequence into a practical specification test based on equalities of distribution functions. We identify complier marginal outcome distributions using Abadie-style weights and estimate them with cross-fitting when covariates are present. We enforce valid CDF shape by monotone rearrangement and truncation. We then test the identity with a uniform Kolmogorov–Smirnov or Cramér–von Mises statistic and a complementary overidentified GMM moment test with robust covariance. We also give a falsification placebo based on predicted noncompliance. Simulations show correct size and useful power against exclusion violations and defiers. Code and replication tables are provided.

1 Introduction

Many randomized encouragement designs have imperfect compliance. Only a fraction of units take up treatment when encouraged. Examples include phone-bank get-out-the-vote campaigns and draft lotteries. Instrumental variables identify the local average treatment effect under three conditions: random assignment, exclusion, and monotonicity [Imbens and Angrist, 1994, Angrist et al., 1996]. Exclusion encodes a simple idea. The instrument changes outcomes only through treatment. In a GOTV setting, assignment that never reaches a person should not shift turnout or downstream outcomes. In a draft lottery, a low number should matter for outcomes only through service.

These conditions imply a sharp distributional consequence. When the instrument Z changes, only compliers change treatment status. As a result, only compliers can move the outcome distribution. Let $F_{Y|Z=z}$ denote the distribution of the observed outcome Y under assignment z. Let p_C be the share of compliers. Let F_{1C} and F_{0C} denote the complier potential-outcome marginals for $Y(1)$ and $Y(0)$. Under the LATE assumptions,

$$F_{Y|Z=1}(y) - F_{Y|Z=0}(y) \;=\; p_C\Big(F_{1C}(y) - F_{0C}(y)\Big) \quad \text{for all } y \in \mathbb{R}.$$

This equality states that the entire population shift in the outcome distribution across assignment equals the treated-versus-control shift among compliers, scaled by the complier share.

This paper develops a specification test of that consequence which can sit alongside conventional IV estimation. We identify the complier marginals using Abadie-style weighting [Abadie, 2002, Abadie, 2003]. We implement estimation with cross-fitting when covariates are present [Chernozhukov et al., 2018] and enforce valid CDF shape with monotone rearrangement [Chernozhukov et al., 2009]. We then test the identity over the full outcome support using uniform Kolmogorov–Smirnov or Cramér–von Mises statistics and a complementary GMM moment test with robust covariance. We also provide a falsification placebo that targets regions with near-zero predicted compliance.

*See https://github.com/finite-sample/late_iv

2 Related work

Three strands are closest and clarify how our approach differs.

Distributional tests among compliers. Abadie identifies and estimates complier outcome distributions and provides bootstrap tests for distributional hypotheses within the complier group [Abadie, 2002, Abadie, 2003, Abadie et al., 2002]. Those tests ask whether treatment shifts the complier distribution. We answer a different question. We reconcile the population CDF shift across assignment with the complier treated-versus-control shift scaled by the complier share. We therefore test a global identity that must hold under the LATE conditions rather than a within-compliers difference.

Inequality-based validity tests. Huber and Mellace derive inequality moment constraints implied by the LATE assumptions and provide tests based on those inequalities. Laffers studies sharpness of such implications [Huber and Mellace, 2015, Laffers, 2017]. Our approach targets a different implication. We test an equality of distribution functions over the full outcome support.

Exogeneity tests. Kitagawa proposes a specification test for instrument validity that exploits nonnegativity implications for complier densities and uses a variance-weighted Kolmogorov–Smirnov statistic. Mourifié and Wan develop tests for LATE assumptions [Kitagawa, 2015, Mourifié and Wan, 2017]. Our test compares the observed CDF shift across assignment to the shift implied by complier marginals and the complier share. The two approaches are complementary.

3 Setup and consequence

Units are $i = 1, \dots, n$. The instrument is $Z \in \{0, 1\}$. Treatment is $D \in \{0, 1\}$. The outcome is $Y \in \mathbb{R}$. Potential treatment is $D(z)$. Potential outcomes are $Y(d)$.

The analysis uses four conditions. Assignment is random. Exclusion holds so that assignment affects outcomes only through treatment. Monotonicity holds so that there are no defiers. SUTVA holds.

Define the complier share

$$p_C \equiv \Pr\{D(1) > D(0)\} \;=\; \mathbb{E}[D \mid Z{=}1] - \mathbb{E}[D \mid Z{=}0].$$

Let $F_{Y|Z=z}$ be the distribution of Y when $Z = z$. Let F_{1C} and F_{0C} denote the complier potential-outcome marginals for $Y(1)$ and $Y(0)$.

Proposition 1 (Only compliers move)**.** *Under random assignment, exclusion, and monotonicity, the following equality holds for all real y:*

$$F_{Y|Z=1}(y) - F_{Y|Z=0}(y) \;=\; p_C\big(F_{1C}(y) - F_{0C}(y)\big).$$

The proof decomposes the outcome distributions by principal strata and uses exclusion to remove terms for always takers and never takers. The complier term remains and is scaled by p_C.

4 Identification and estimation

The complier marginals are identified without joint rank assumptions. Let X denote observed covariates. Define

$$e(X) = \Pr(Z = 1 \mid X), \qquad p_z(X) = \mathbb{E}[D \mid Z = z, X], \qquad p_C = \mathbb{E}\big[p_1(X) - p_0(X)\big].$$

For any measurable function g,

$$\mathbb{E}\big[g(Y(1)) \mid C\big] \;=\; \frac{\mathbb{E}\Big[g(Y)\,\frac{Z}{e(X)}\big(D - p_0(X)\big)\Big]}{p_C}, \qquad \mathbb{E}\big[g(Y(0)) \mid C\big] \;=\; \frac{\mathbb{E}\Big[g(Y)\,\frac{1-Z}{1-e(X)}\big(p_1(X) - D\big)\Big]}{p_C}.$$

Choosing indicator kernels $g_y(u) = \mathbf{1}\{u \le y\}$ yields the complier CDFs F_{1C} and F_{0C}. In a pure randomized trial without covariates, $e \equiv \frac{1}{2}$ and p_z reduces to group means.

Estimated complier CDFs must be valid distribution functions. After computing the weighted CDF curves, apply monotone rearrangement on a fine grid and truncate to the unit interval [Chernozhukov et al., 2009]. This step enforces monotonicity and proper range and affects finite-sample shape only.

All first-stage nuisance functions are estimated with cross-fitting when covariates are present [Chernozhukov et al., 2018]. Nuisance functions are fit on training folds and applied to held-out folds. This stabilizes inference for the weighted empirical processes that define the complier CDFs.

5 Two tests of the consequence

5.1 Uniform distributional test

Define the estimated difference

$$\hat{\Delta}(y) = \hat{F}_{Y|Z=1}(y) - \hat{F}_{Y|Z=0}(y) - \hat{p}_C\big(\hat{F}_{1C}(y) - \hat{F}_{0C}(y)\big).$$

Test the null that $\hat{\Delta}$ equals zero for all y. Use either the Kolmogorov–Smirnov statistic $T_\infty = \sup_y |\hat{\Delta}(y)|$ or the Cramér–von Mises statistic $T_2 = \int \hat{\Delta}(y)^2\, d\hat{H}(y)$, where $\hat{H}$ is a pooled empirical measure on a fine grid. With covariates, obtain critical values using a multiplier bootstrap that respects cross-fitting. Generate mean-zero unit-variance multipliers and reweight held-out contributions while keeping fold-specific nuisance functions fixed. The bootstrap delivers critical values for the uniform statistics.

5.2 GMM moment test

Proposition 1 implies a family of unconditional moments. For any bounded function g,

$$\mathbb{E}[g(Y) \mid Z{=}1] - \mathbb{E}[g(Y) \mid Z{=}0] \;-\; p_C\big(\mathbb{E}[g(Y(1)) \mid C] - \mathbb{E}[g(Y(0)) \mid C]\big) \;=\; 0.$$

Select a dictionary of indicator functions $g_j(y) = \mathbf{1}\{y \le t_j\}$ at prespecified cutpoints $\{t_j\}_{j=1}^J$. Stack the moments into a vector m. Use the robust GMM J-test

$$T_J \;=\; n\, m^\top \hat{W}^{-1} m$$

with a heteroskedasticity-robust sandwich covariance $\hat{W}$. This test targets specific regions of the distribution when the cutpoints are chosen in the tails.

5.3 Placebo falsification with predicted noncompliance

Define the compliance propensity $c(X) = p_1(X) - p_0(X)$. In regions where $c(X)$ is close to zero, the instrument should have no effect on outcomes if the LATE conditions hold. Estimate $c(X)$ with cross-fitting. Define a smoothed weight $w_\tau(X) = K\big((\hat{c}(X) - \tau)/h\big)$ for a kernel K and bandwidth h. Test that the conditional mean difference of Y across Z equals zero when averaged with weights $w_\tau(X)$. Construct a max test over a grid of thresholds and use a multiplier bootstrap to obtain critical values. This falsification probes exclusion in parts of the covariate space that have near-zero compliance.

6 Inference guidance

Report the complier share $\hat{p}_C$. When $\hat{p}_C$ is very small, the variance of the weighted complier estimators is large and the uniform test may have low power. In that case emphasize the placebo falsification and report randomization checks. In trials with known assignment, randomization inference is attractive. In observational encouragement designs rely on cross-fitting and the multiplier bootstrap.

7 Simulation study

The simulation study evaluates size and power for the uniform Cramér–von Mises test and for the GMM moment test with ten indicator cutpoints. The first block focuses on a randomized trial without covariates. The second block adds covariates and cross-fitted nuisance functions.

7.1 Design without covariates

Principal strata are generated with shares that sum to one. Assignment is Bernoulli with rate one half. Potential treatment equals type rules. The baseline outcome $Y(0)$ is standard normal. Complier treatment effects are normal with mean 0.5 and standard deviation 0.4. Observed outcomes equal $Y(D)$.

The study considers a null that satisfies LATE. It then considers exclusion violations that add a direct effect γZ to Y for noncompliers. It also considers monotonicity violations created by a positive mass of defiers. The sample size is two thousand unless noted. Critical values are estimated under the null by Monte Carlo and then held fixed to evaluate size and power.

Table 1: Size and power at five percent. Two hundred fifty replications per row. Critical values estimated from four hundred null draws.

Scenario	p_C	Alt	T_2 mean	T_2 SD	Reject T_2	Reject T_J
Null	0.30	none	0.000571	0.000543	0.056	0.056
Null	0.10	none	0.000795	0.000665	0.056	0.052
Exclusion $\gamma = 0.2$	0.30	excl	0.000779	0.000593	0.100	0.092
Exclusion $\gamma = 0.5$	0.30	excl	0.007091	0.002006	1.000	1.000
Defiers five percent	0.30	def	0.000968	0.000690	0.140	0.132
Defiers ten percent	0.30	def	0.001470	0.000943	0.312	0.336

Two further checks vary sample size and complier share.

Table 2: Sensitivity to sample size and complier share.

Scenario	n	p_C	T_2 mean	T_2 SD	Reject T_2	Reject T_J
Exclusion $\gamma = 0.2$	1000	0.30	0.000992	0.000913	0.060	0.064
Exclusion $\gamma = 0.2$	4000	0.30	0.000580	0.000363	0.084	0.088
Defiers five percent	2000	0.10	0.001518	0.001014	0.180	0.176
Defiers five percent	2000	0.50	0.000562	0.000552	0.080	0.076

The uniform test holds size near five percent under the null. Power rises with the strength of the violation. Exclusion effects of moderate size are detected reliably at larger n. The tests also detect monotonicity violations and become more powerful as the mass of defiers grows.

7.2 Design with covariates

The covariate design adds a vector X that shifts both compliance and potential outcomes. The conditional compliance rate $p_1(X) - p_0(X)$ depends on X through a logistic index. The baseline outcome $Y(0)$ depends on X through a linear index and a sine term. The complier effect distribution is the same as above. All first-stage functions are estimated with cross-fitting.

The uniform test uses a multiplier bootstrap with fold-locked nuisance functions. The placebo falsification uses a kernel weight on predicted noncompliance and a max test over a grid of thresholds.

The results show size near five percent under the conditional LATE null. The uniform test and the GMM test detect exclusion violations that create conditional mean differences in regions with near-zero compliance. The placebo falsification has high power in those regions. Detailed tables and code are included in the replication files.

8 Practice recommendations

Report the complier share. When it is small, emphasize falsification in low-compliance regions and randomization checks. Enforce valid shape for the complier CDFs through rearrangement and truncation. Use cross-fitting for all nuisance estimates. Use robust covariance in the GMM test. Prefer the uniform test when the research question is global movement of the outcome distribution. Use the GMM test to probe specific parts of the distribution. Visualize both sides of the identity by plotting the two CDF differences with uniform bands.

9 Conclusion

The LATE assumptions imply a simple and powerful consequence: only compliers move when the instrument flips. The paper provides identification, estimators, and tests that turn this consequence into a practical specification check. The tests complement standard IV estimation and give a distributional view that distinguishes between compliance-driven movement and violations such as exclusion failures and defiers.

References

[Abadie, 2002] Abadie, A. (2002). Bootstrap tests for distributional treatment effects in instrumental variable models. *Journal of the American Statistical Association*, 97(457):284–292.

[Abadie, 2003] Abadie, A. (2003). Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics*, 113(2):231–263.

[Abadie et al., 2002] Abadie, A., Angrist, J. D., and Imbens, G. W. (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica*, 70(1):91–117.

[Angrist et al., 1996] Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455.

[Chernozhukov et al., 2018] Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68.

[Chernozhukov et al., 2009] Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika*, 96(3):559–575.

[Huber and Mellace, 2015] Huber, M. and Mellace, G. (2015). Testing instrument validity for late identification based on inequality moment constraints. *Review of Economics and Statistics*, 97(2):398–411.

[Imbens and Angrist, 1994] Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475.

[Kitagawa, 2015] Kitagawa, T. (2015). A test for instrument validity. *Econometrica*, 83(5):2043–2063.

[Laffers, 2017] Laffers, L. (2017). A note on testing instrument validity for the identification of late. *Empirical Economics*, 53(3):1123–1131.

[Mourifié and Wan, 2017] Mourifié, I. and Wan, Y.-J. (2017). Testing local average treatment effect assumptions. *Review of Economics and Statistics*, 99(2):305–313.