KERNELIZED IV CONFOUNDING SCREEN FOR TESTING CONDITIONAL EXOGENEITY

Anonymous authors
Paper under double-blind review

ABSTRACT

Instrumental variables (IVs) provide a powerful route to causal identification when latent confounding is present, and conditional exogeneity is one of their most basic credibility requirements. Existing methods for testing IV conditional exogeneity typically rely on structural restrictions, such as linearity or discrete-support conditions, or require auxiliary side information, such as assignment mechanisms and negative controls. We therefore propose Kernelized IV Confounding Screen (KICS) for testing IV conditional exogeneity under more flexible structural conditions and without requiring auxiliary side information. The key idea of KICS is to compare the coefficient systems of an ordinary kernel regression and a candidate-IV-induced weighted kernel regression after removing the part of the candidate IV explained by observed covariates. Theoretically, at the population level, the two coefficient systems coincide under the null hypothesis and differ under detectable alternatives; the estimated coefficient discrepancy is asymptotically normal, enabling a Wald-type test for practical inference. Experiments in synthetic and real-data-based settings show that KICS detects conditional exogeneity violations in flexible nonlinear settings, remains competitive in settings favorable to specialized baselines, and yields real-data results consistent with prior IV assessments. Overall, KICS provides a practical approach to IV conditional exogeneity testing in settings where existing methods are difficult to apply.

1 INTRODUCTION

Causal effect estimation is central to decision-making in high-stakes domains such as healthcare, public policy, and finance (Prosperi et al., 2020; Dahabreh & Bibbins-Domingo, 2024; Matthay & Glymour, 2022; Kumar et al., 2025). In such settings, decisions often depend on credible causal evidence. Randomized controlled trials (RCTs) are widely regarded as the gold standard for causal inference, but they are often limited by ethical constraints, implementation difficulty, and cost. As a result, observational studies are frequently used in practice as a more feasible source of evidence (Pearl, 2009; Hernán & Robins, 2020). However, causal conclusions drawn from observational data are inherently vulnerable to latent confounding, making the credibility of observational causal inference fundamentally difficult to assess.

Instrumental variables (IVs) offer a classical strategy for addressing latent confounding. Rather than requiring all confounders to be observed, IV methods rely on variables that affect the treatment, influence the outcome only through the treatment, and are otherwise independent of outcome-side latent confounding after conditioning on observed covariates (Angrist et al., 1996; Labrecque & Swanson, 2018). These requirements are commonly summarized as three core IV assumptions: relevance, the exclusion restriction, and conditional exogeneity. This idea is powerful, but also fragile: the credibility of an IV depends on whether it truly satisfies all three assumptions.

Assessing IV credibility is therefore indispensable before relying on IV-based identification. A valid IV can mitigate latent confounding, whereas an invalid IV may transmit latent bias into causal effect estimates (Angrist et al., 1996; Conley et al., 2012). Since candidate IVs are often justified by domain knowledge rather than guaranteed by design, their validity cannot be assumed (Hernán & Robins, 2020). IV validity comprises distinct requirements: relevance and exclusion-related restrictions have received extensive diagnostic attention (Keane & Neal, 2024; Lewis & Mertens, 2025), whereas direct tests of **conditional exogeneity**, a key requirement for IV validity, remain comparatively limited,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

especially for a single candidate IV in flexible observational settings. Specifically, conditional exogeneity requires that, once observed covariates are accounted for, any remaining variation in the candidate IV should be unrelated to latent factors that also affect the outcome.

Existing methods for testing or falsifying IV conditional exogeneity can be broadly grouped into two classes. One class obtains testability through structural restrictions, such as discrete-support or linear assumptions (Kitagawa, 2015; Sun, 2023; Burauel, 2023; Xie et al., 2022; Wiedermann & Shi, 2026). These methods can be effective when their assumptions are credible, but in many practical settings, continuous variables, nonlinear relationships, and more complex data structures make the required structural restrictions difficult to satisfy or justify. A second class leverages auxiliary side information, such as assignment mechanisms, negative controls, or auxiliary variables (Branson & Keele, 2020; Danieli et al., 2026; Guo et al., 2026). These methods can also be powerful, but require additional information that may be unavailable or difficult to justify in observational studies. Taken together, existing approaches typically rely on either structural restrictions or auxiliary side information, which raises an important open question:

Can IV conditional exogeneity be tested under more flexible structural conditions without requiring auxiliary side information?

To address this gap, we propose **KICS** (**K**ernelized **I**V **C**onfounding **S**creen), a kernel-based procedure for testing IV conditional exogeneity. KICS is based on the following principle: if a candidate IV is conditionally exogenous, then reweighting samples using its residual variation after removing the part explained by observed covariates should not systematically change the regression structure of the outcome given treatment and covariates. Conversely, if the candidate IV remains associated with outcome-side latent confounding, such reweighting can induce a systematic coefficient discrepancy. Operationally, KICS proceeds in three steps: it residualizes the candidate IV with respect to observed covariates, fits an ordinary kernel regression of the outcome on treatment and covariates, and compares it with a candidate-IV-induced weighted kernel regression. The resulting coefficient discrepancy yields a Wald-type test for conditional exogeneity. The main contributions of this paper are as follows:

- We propose KICS for testing the conditional exogeneity of a single candidate IV under more flexible structural conditions than existing tests and without requiring auxiliary side information, particularly in nonlinear settings with continuous or otherwise complex variables.
- We establish the theoretical foundation of KICS, including population-level coefficient equivalence under the null, coefficient discrepancy under detectable alternatives, and asymptotic normality leading to a Wald-type test.
- We evaluate KICS across synthetic and real-data-based settings, showing that it performs effectively in flexible nonlinear settings, remains competitive under the assumptions required by specialized baselines, and yields real-data results consistent with prior IV assessments.

2 Related Work

2.1 Diagnostics for Relevance and Exclusion-Related Restrictions

Instrumental variables (IVs) are widely used for causal identification in the presence of latent confounding. A valid IV requires three core conditions: relevance, the exclusion restriction, and conditional exogeneity. Existing diagnostic methods extensively address relevance and exclusion-related restrictions as components of IV credibility. Relevance diagnostics examine instrument strength and weak-IV behavior (Stock & Yogo, 2005; Angrist & Pischke, 2009; Keane & Neal, 2024), while exclusion-related concerns are commonly addressed through sensitivity and falsification analyses (Conley et al., 2012; Angrist et al., 1996; Lewis & Mertens, 2025). KICS complements these efforts by directly testing the conditional exogeneity of a single candidate IV.

2.2 Testing IV Conditional Exogeneity

Existing tests of IV conditional exogeneity typically obtain testability through either structural restrictions or auxiliary side information. One class derives observable implications from structural

restrictions, particularly discrete-support or linear assumptions. In discrete-support settings, restrictions on the joint distribution of treatment, outcome, and instrument yield instrumental inequalities and related nonparametric characterizations (Pearl, 2013; Balke & Pearl, 1997; Bonet, 2013; Kédagni & Mourifié, 2020); representative tests include Kitagawa's test for binary treatments and discrete instruments (Kitagawa, 2015) and Sun's test for multivalued treatments (Sun, 2023). Another line relies on linear assumptions in continuous-IV settings (Gunsilius, 2021; Wiedermann & Shi, 2026; Burauel, 2023; Xie et al., 2022). These approaches provide principled falsification tools, but their applicability depends on whether the required structural restrictions are credible in practice.

A second class leverages auxiliary side information beyond the observed variables. Randomization-based approaches assess candidate-IV consistency with a specified assignment mechanism by comparing observed balance or bias with its implied distribution (Branson & Keele, 2020; Keele et al., 2019; Pizer, 2016). Negative-control methods introduce additional variables to reveal residual threats to identification when the IV is invalid (Lipsitch et al., 2010; Arnold et al., 2016; Shi et al., 2020; Yang et al., 2024; Penning de Vries & Groenwold, 2023; Danieli et al., 2026). More recently, AIT derives testable implications for IV validity through auxiliary constructions in additive nonlinear models (Guo et al., 2026). These methods can be powerful, but their applicability depends on the availability and credibility of the required auxiliary information.

KICS instead tests IV conditional exogeneity under more flexible structural conditions and without requiring auxiliary side information. This broadens applicability to settings where existing specialized tests are difficult to use.

2.3 Kernel-Based Confounding Detection

Kernel methods provide flexible RKHS representations of nonlinear relationships when simple parametric forms may be inadequate (Aronszajn, 1950; Kimeldorf & Wahba, 1971). They are widely used for distribution and independence testing (Gretton et al., 2012; Muandet et al., 2017; Zhang et al., 2012; Pfister et al., 2018), adjustment for observed or proxy confounding (Hazlett, 2020; Mastouri et al., 2021), and nonlinear IV estimation (Singh et al., 2019). A separate diagnostic line detects latent confounding through instability between ordinary and reweighted regressions; in linear settings, this appears in higher-order least squares and related diagnostics (Schultheiss et al., 2024), while KRCD extends the principle to nonlinear single-environment settings through RKHS coefficient discrepancy (Chen et al., 2026). KICS draws on this regression-discrepancy principle but develops an IV-specific formulation for conditional exogeneity testing, constructing candidate-IV-induced reweighting from residual IV variation and using the resulting coefficient instability as the testing signal.

3 Preliminaries and Problem Setup

3.1 Problem Setup

Instrumental variables. Instrumental variables (IVs) provide a classical strategy for causal identification under latent confounding. A valid IV requires relevance, exclusion restriction, and conditional exogeneity: the IV must affect the treatment, affect the outcome only through the treatment, and remain independent of outcome-side latent confounding after conditioning on observed covariates, respectively. In this paper, we focus on testing the conditional exogeneity of a candidate IV.

Candidate IV and target hypothesis. We consider an observational setting in which only the outcome Y, treatment T, observed covariates X, and a candidate IV Z are available. Let U_Y denote unobserved factors that affect the outcome and may threaten causal identification. We study whether the candidate IV Z satisfies

$$H_0 : Z \perp U_Y \mid X.$$

That is, after conditioning on X, the candidate IV should no longer be associated with outcome-side latent confounding. If this condition is violated, residual variation in Z can remain associated with latent factors that affect Y. Consequently, the treatment variation induced by Z may carry latent-confounding signal, compromising IV identification.

Challenges. Existing approaches do not directly cover this general setting, as IV conditional-exogeneity tests typically rely on structural restrictions or auxiliary side information. Table 1 summarizes these applicability differences across representative regimes. In contrast, we consider the

Table 1: Applicability of KICS and baselines across representative regimes.

Method	General Nonlinear	Linear	Discrete	Aux-info
KICS (Ours)	✓	✓	✓	✓
LinNG-IV (Wiedermann & Shi, 2026)	✗	✓	✗	✗
IV-PIM (Burauel, 2023)	✗	✓	✗	✗
IV-GIN (Xie et al., 2022)	✗	✓	✗	✗
Kitagawa (Kitagawa, 2015)	✗	✗	✓	✗
Sun (Sun, 2023)	✗	✗	✓	✗
Branson (Branson & Keele, 2020)	✗	✗	✗	✓
NCF (Danieli et al., 2026)	✗	✗	✗	✓
AIT (Guo et al., 2026)	✗	✗	✗	✓

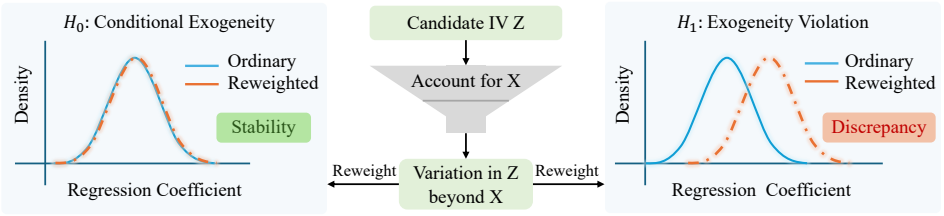

Figure 1: Intuition of Candidate-IV Reweighting for Conditional Exogeneity Testing.

setting where only the observed quadruple (Y, T, X, Z) is available and no such additional ingredients are assumed. The challenge is therefore to test $Z \perp U_Y \mid X$ using only observable data, even though the relevant latent confounding is unobserved.

3.2 Regression-Comparison Principle for IV Conditional Exogeneity

KICS builds on a regression-comparison principle for detecting latent-confounding-related instability: under the relevant exogeneity condition, an appropriate reweighting should leave the outcome regression structure stable, whereas reweighting by a variable associated with outcome-side latent confounding can induce coefficient instability.

Higher-order regression as a diagnostic principle. Higher-order regression operationalizes this idea by comparing ordinary and reweighted regressions. In nonlinear settings, KRCD implements this comparison in a finite-dimensional RKHS subspace (Chen et al., 2026). Let $V = (T, X)$ and let $\{\phi_i\}_{i=1}^P$ be a basis of the chosen subspace. KRCD compares

$$\boldsymbol{\alpha}_V^{\mathrm{KLS}} = \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^P} \mathbb{E}\left[\left(Y - \sum_{i=1}^P \alpha_i \phi_i(V)\right)^2\right],$$

$$\boldsymbol{\alpha}_V^{\mathrm{HKLS}} = \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^P} \mathbb{E}\left[\left(Y - \sum_{i=1}^P \alpha_i \phi_i(V)\right)^2 \|V\|^2\right].$$

Their discrepancy serves as a signal for latent confounding in the treatment–outcome relationship.

IV intuition. This principle has a natural analogue in IV reasoning. After conditioning on observed covariates, residual variation in a credible IV should be unrelated to outcome-side latent confounding. Thus, under conditional exogeneity, reweighting by this residual variation should not systematically change the regression structure of Y given (T, X). A systematic change suggests that residual IV variation may still carry outcome-side latent confounding signal, as illustrated in Figure 1.

Target-specific reformulation. Adapting this intuition to IV conditional exogeneity testing is nontrivial. KRCD targets latent confounding in the treatment–outcome relationship, whereas the target here is the conditional exogeneity component of IV credibility, $Z \perp U_Y \mid X$. Moreover, KRCD's

higher-order weight is tied to the regression input itself, while IV conditional exogeneity concerns residual IV variation after removing the part explained by X. Thus, the regression-comparison idea must be reformulated with an IV-induced weight and interpreted as a test of conditional exogeneity, not as validation of all IV assumptions.

4 KERNELIZED IV CONFOUNDING SCREEN

We propose **KICS** (**K**ernelized **I**V **C**onfounding **S**creen), a kernel-based procedure for testing conditional exogeneity of a candidate IV under flexible nonlinear structural conditions without requiring auxiliary side information. Operationally, KICS first residualizes the candidate IV with respect to observed covariates, then compares an ordinary kernel regression of the outcome on treatment and covariates with a candidate-IV-induced weighted kernel regression, and finally converts the resulting coefficient discrepancy into a formal hypothesis test for conditional exogeneity.

Residualization as nuisance estimation. Its first step estimates the nuisance function $\mathbb{E}[Z \mid X]$ to isolate residual variation in the candidate IV. This residualization is statistical preprocessing rather than a low-dimensional structural restriction on the causal model, and can be implemented with any sufficiently accurate regression learner.

4.1 THEORETICAL FOUNDATION

To construct KICS, we first isolate the residual variation of the candidate IV by defining

$$R_Z := Z - \mathbb{E}[Z \mid X].$$

Here, $\mathbb{E}[Z \mid X]$ is treated as a nuisance regression estimated from the observed data, so that R_Z captures the component of the candidate IV unexplained by observed covariates. Under conditional exogeneity, this residual variation should no longer carry outcome-side latent confounding signal. KICS uses R_Z^2 to induce sample reweighting.

Leveraging the finite-dimensional RKHS representation introduced in the previous section, we formulate two population estimators in the P-dimensional subspace $\mathcal{S}_P$ spanned by basis functions $\{\phi_i\}_{i=1}^P$, where P denotes the basis size. Let $V = (T, X)$.

The ordinary **Kernelized Least Squares** (KLS) estimator is defined as

$$\boldsymbol{\alpha}_V^{\mathrm{KLS}} = \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^P} \mathbb{E}\left[\left\|Y - \sum_{i=1}^P \alpha_i \phi_i(V)\right\|^2\right], \tag{1}$$

and the **IV-weighted Kernelized Least Squares** (IVW-KLS) estimator is defined as

$$\boldsymbol{\alpha}_V^{\mathrm{IVW}} = \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^P} \mathbb{E}\left[\left\|Y - \sum_{i=1}^P \alpha_i \phi_i(V)\right\|^2 R_Z^2\right]. \tag{2}$$

Sufficient conditions for coefficient equivalence. The population coefficient comparison is justified under sufficient conditions: the chosen RKHS subspace represents the conditional outcome regression $g_\star(V) := \mathbb{E}[Y \mid V]$; under H_0, latent-weight interaction and noise-orthogonality conditions imply a vanishing weighted residual moment; and under detectable alternatives, this weighted moment is nonzero. These conditions ensure coefficient equality under conditional exogeneity and coefficient discrepancy under detectable alternatives; details are given in Appendix A.1.2.

Core principle of KICS. Based on equation 1 and equation 2, KICS tests IV conditional exogeneity by comparing $\boldsymbol{\alpha}_V^{\mathrm{KLS}}$ and $\boldsymbol{\alpha}_V^{\mathrm{IVW}}$ coordinatewise. Under conditional exogeneity, reweighting by the residual variation of the candidate IV should not systematically alter the regression structure of Y on (T, X), whereas violations can induce coefficient instability. This principle is formalized in the following theorem.

Theorem 1 (Population coefficient equivalence under conditional exogeneity) *Under the sufficient conditions in Appendix A.1.2, the following holds:*

$$H_0 : \forall i \in [P],\ \boldsymbol{\alpha}_V^{\mathrm{KLS}}[i] = \boldsymbol{\alpha}_V^{\mathrm{IVW}}[i],$$
$$H_1 : \exists i \in [P],\ \boldsymbol{\alpha}_V^{\mathrm{KLS}}[i] \neq \boldsymbol{\alpha}_V^{\mathrm{IVW}}[i].$$

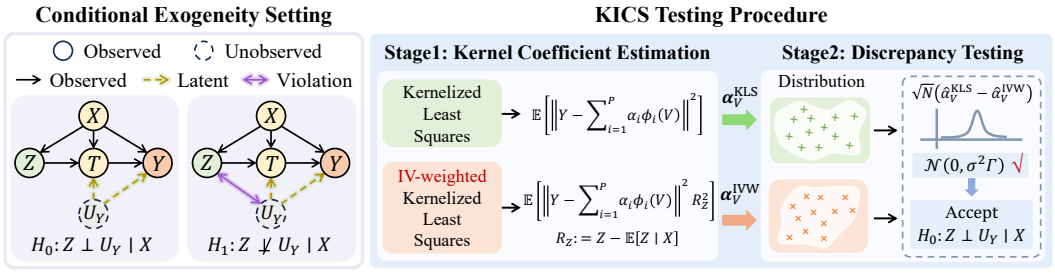

Figure 2: Overview of KICS for IV Conditional Exogeneity Testing.

The proof of Theorem 1 is provided in Appendix A.1.3. The theorem characterizes, at the population level, the equality of the KLS and IVW-KLS coefficient systems under the null and their discrepancy under detectable alternatives.

4.2 Hypothesis Testing with Finite Samples

Although Theorem 1 establishes this result at the population level, the estimators in equation 1 and equation 2 are not directly observable in finite samples. Therefore, in practical finite-sample settings, KICS is implemented through empirical counterparts of the two coefficient systems, with inference based on their estimated discrepancy.

To this end, let $\widehat{m}(X_i) \approx \widehat{\mathbb{E}}[Z \mid X_i]$ denote a first-stage estimate of the nuisance regression $\mathbb{E}[Z \mid X]$, and define the empirical residualized candidate IV as $\hat{R}_{Z,i} = Z_i - \widehat{m}(X_i)$. Let $\widehat{\Psi}_Z = \mathrm{diag}(\hat{R}^2_{Z,1}, \ldots, \hat{R}^2_{Z,N})$ be the induced diagonal weight matrix. We further write $K \in \mathbb{R}^{P\times N}$ for the kernel design matrix induced by the chosen P basis functions on the regression input $V = (T, X)$, and let $\boldsymbol{Y} = [Y_1, \ldots, Y_N]^\top$ denote the outcome vector. Under ridge regularization with parameter $\xi > 0$, the empirical KLS and IVW-KLS estimators admit the following closed-form solutions.

Theorem 2 (Finite-sample KICS estimators) *The empirical KLS and IVW-KLS estimators are*

$$\widehat{\boldsymbol{\alpha}}_V^{\mathrm{KLS}} = (KK^\top + N\xi I_P)^{-1} K\boldsymbol{Y}, \tag{3}$$

$$\widehat{\boldsymbol{\alpha}}_V^{\mathrm{IVW}} = (K\widehat{\Psi}_Z K^\top + N\xi I_P)^{-1} K\widehat{\Psi}_Z \boldsymbol{Y}. \tag{4}$$

The proof is given in Appendix A.2. In finite samples, the difference between $\widehat{\boldsymbol{\alpha}}_V^{\mathrm{IVW}}$ and $\widehat{\boldsymbol{\alpha}}_V^{\mathrm{KLS}}$ is perturbed by estimation error in the residualized candidate IV, sampling variability, and outcome noise. Therefore, practical hypothesis testing requires an asymptotic characterization of this estimated coefficient discrepancy. The regularity conditions for asymptotic inference are stated in Appendix A.3.

Theorem 3 (Asymptotic distribution of the estimated coefficient discrepancy) *Assume the regularity conditions hold. Under the conditional exogeneity null hypothesis H_0, as $N \to \infty$,*

$$\sqrt{N}\left(\widehat{\boldsymbol{\alpha}}_V^{\mathrm{IVW}} - \widehat{\boldsymbol{\alpha}}_V^{\mathrm{KLS}}\right) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \Sigma_Z), \tag{5}$$

where the asymptotic covariance matrix Σ_Z is consistently estimable.

The proof and Wald-test details are provided in Appendix A.3. Theorem 3 provides the asymptotic basis for implementing a Wald-type test in finite samples based on $\widehat{\boldsymbol{\alpha}}_V^{\mathrm{IVW}} - \widehat{\boldsymbol{\alpha}}_V^{\mathrm{KLS}}$, thereby enabling practical inference on IV conditional exogeneity.

4.3 Complexity Analysis

We derive the overall complexity of KICS as $O\big(C_m(N, d_X) + PNd_V + P^2N + P^3\big)$, where N is the sample size, P is the basis size, $d_V = \dim(V)$ is the dimension of the regression input $V = (T, X)$, and $C_m(N, d_X)$ denotes the cost of estimating the residualization model for the candidate IV. Detailed derivations are provided in Appendix A.4.

Algorithm 1 Kernelized IV Confounding Screen (KICS)

Input: Outcome $Y \in \mathbb{R}^N$, treatment $T \in \mathbb{R}^N$, observed covariates $X \in \mathbb{R}^{N \times d_X}$, candidate IV $Z \in \mathbb{R}^N$, basis size P, regularization parameter $\lambda = N\xi$, significance level α.
Output: Test decision for $H_0 : Z \perp U_Y \mid X$

1: Set $V_i \leftarrow (T_i, X_i)$ and estimate $\widehat{m}(X_i) \approx \mathbb{E}[Z \mid X_i]$
2: Compute $\hat{R}_{Z,i} \leftarrow Z_i - \widehat{m}(X_i)$ and $\widehat{\Psi}_Z \leftarrow \mathrm{diag}(\hat{R}^2_{Z,1}, \ldots, \hat{R}^2_{Z,N})$
3: Construct the kernel design matrix $K \in \mathbb{R}^{P \times N}$ on $\{V_i\}_{i=1}^N$
4: $\widehat{\boldsymbol{\alpha}}_V^{\mathrm{KLS}} \leftarrow (KK^\top + \lambda I_P)^{-1} K\boldsymbol{Y}$
5: $\widehat{\boldsymbol{\alpha}}_V^{\mathrm{IVW}} \leftarrow (K\widehat{\Psi}_Z K^\top + \lambda I_P)^{-1} K\widehat{\Psi}_Z\boldsymbol{Y}$
6: Set $\widehat{\Delta} \leftarrow \widehat{\boldsymbol{\alpha}}_V^{\mathrm{IVW}} - \widehat{\boldsymbol{\alpha}}_V^{\mathrm{KLS}}$ and compute a consistent estimate $\widehat{\Sigma}_Z$
7: Compute $Q_N \leftarrow N\widehat{\Delta}^\top \widehat{\Sigma}_Z^{-1} \widehat{\Delta}$
8: **if** $Q_N > \chi^2_{P,1-\alpha}$ **then**
9: **return** Reject H_0
10: **else**
11: **return** Fail to reject H_0
12: **end if**

5 Experiments

Benchmarks. We evaluate KICS across synthetic regimes, including a general nonlinear setting and three specialized regimes: linear, discrete-support, and auxiliary-information. We further consider the Twins semi-synthetic benchmark (Almond et al., 2005) and a real-world Angrist–Evans 1980 U.S. Census application (Angrist & Evans, 1998). The significance level is set to $\alpha = 0.05$ throughout. Further experimental details are provided in Appendix B.1.

Baselines. We compare KICS with representative methods from two classes. The first class requires structural restrictions: LinNG-IV (Wiedermann & Shi, 2026), IV-PIM (Burauel, 2023), IV-GIN (Xie et al., 2022), Kitagawa (Kitagawa, 2015), and Sun (Sun, 2023). The second class requires auxiliary side information: Branson (Branson & Keele, 2020), NCF (Danieli et al., 2026), and AIT (Guo et al., 2026). Full baseline descriptions are deferred to Appendix B.2.

Evaluation Protocols. We report the following three evaluation metrics:

(i) Detection Rate: The proportion of trials in which the method rejects the null hypothesis, reflecting detection performance under violations and false-rejection behavior under the null.

(ii) Null-adjusted Detection Rate: The detection rate at a given confounding strength minus that under zero confounding, measuring the increase in rejection beyond the rate observed under the null.

(iii) ROC Curve & AUC: The ROC curve characterizes performance across thresholds, while AUC summarizes threshold-free discriminative performance.

Questions. The empirical evaluation addresses the following four research questions:

(i) In nonlinear settings without auxiliary side information, can KICS effectively detect candidate IVs that violate conditional exogeneity?

(ii) When structural restrictions or auxiliary side information are available, does KICS remain competitive with baselines tailored to those regimes?

(iii) How does KICS perform in real-data-based experimental settings?

(iv) How sensitive is KICS to sample size and key hyperparameters?

5.1 Detection in Nonlinear Settings without Auxiliary Side Information

Figure 3a evaluates KICS in a general nonlinear continuous setting where specialized structural assumptions do not hold and auxiliary side information is unavailable. KICS's null-adjusted detection rate rises rapidly with confounding strength and remains high under moderate and strong violations, whereas most baselines stay near their null performance or exhibit non-monotonic behavior, including

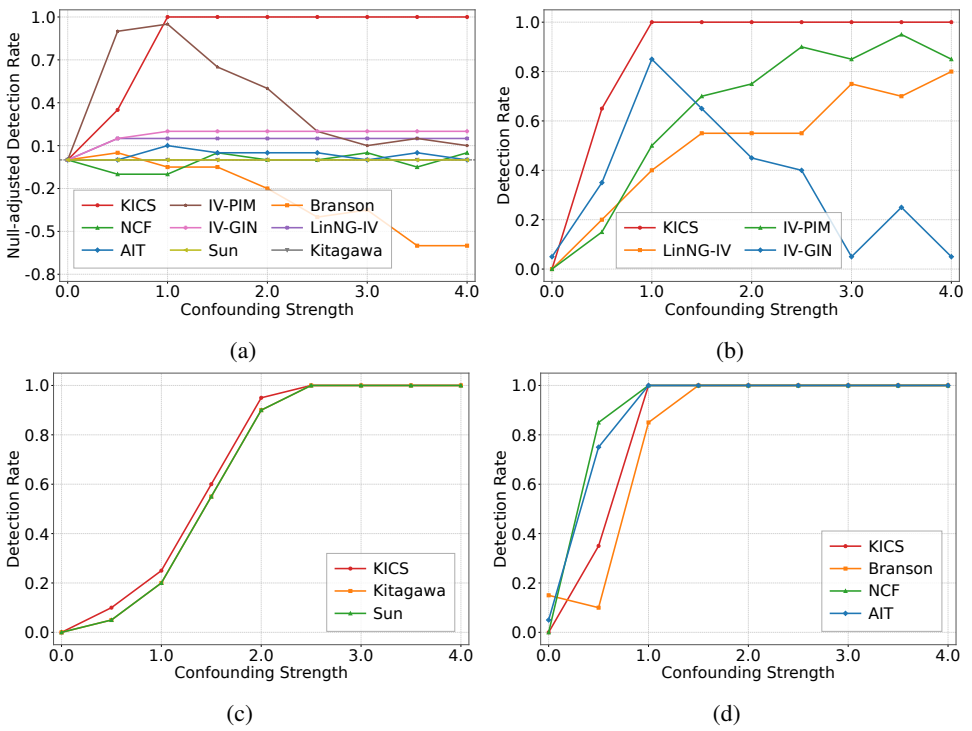

Figure 3: Detection performance across synthetic regimes—(a) Null-adjusted detection rate in the general nonlinear setting. (b) Detection rate in the linear setting. (c) Detection rate in the discrete-support setting. (d) Detection rate in the auxiliary-information setting.

Table 2: Real-world KICS screening in the Angrist–Evans 1980 U.S. Census application.

Candidate IV	**Literature Status**	p**-value**	**Decision**	**Consistency**
Same-sex first two children	Design-supported IV (Angrist & Evans, 1998)	0.8974	Fail to reject H_0	✓
Second-parity twin birth	Exogeneity-challenged IV (Farbmacher et al., 2018)	$< 10^{-4}$	Reject H_0	✓

negative adjusted rates. Overall, KICS effectively detects conditional-exogeneity violations in this challenging setting where specialized baselines are mismatched.

5.2 Competitiveness When Structural Restrictions or Auxiliary Side Information Are Available

Across the three matched regimes, KICS remains competitive without exploiting the specialized conditions available to the baselines. In the linear setting (Figure 3b), KICS achieves strong detection comparable to the linear-specific methods. In the discrete-support setting (Figure 3c), KICS slightly outperforms the specialized tests under weak-to-moderate violations and matches them as confounding strengthens. In the auxiliary-information setting (Figure 3d), specialized methods lead under weak confounding, but KICS quickly reaches comparable high detection. This indicates that broader applicability need not sacrifice detection performance. Overall, KICS retains strong performance even in regimes tailored to specialized baselines.

5.3 Performance in Real-Data-Based Experimental Settings

We further evaluate KICS on the Twins semi-synthetic benchmark and the Angrist–Evans Census application. On Twins, Figures 4a and 4b show clear discrimination and stronger detection as confounding increases, with particularly high performance under moderate and strong violations.

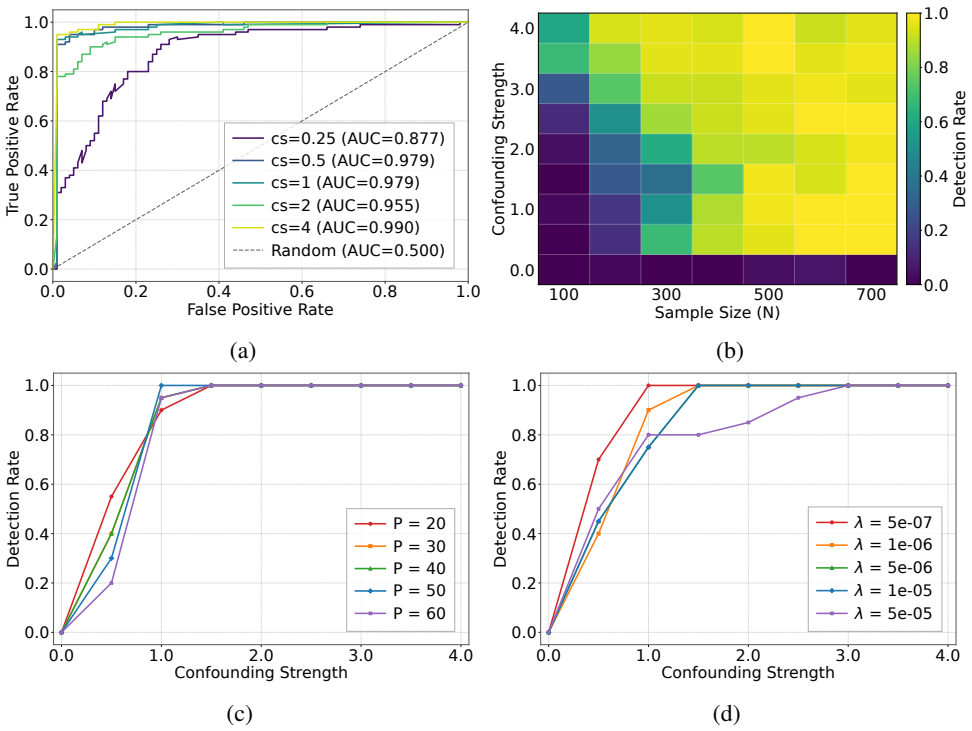

Figure 4: Semi-synthetic performance on Twins and sensitivity analysis of KICS—(a) ROC curves across confounding strengths on Twins. (b) Detection rates across sample sizes and confounding strengths on Twins. (c) Sensitivity to basis size P. (d) Sensitivity to regularization parameter λ.

Because conditional-exogeneity ground truth is unavailable in real-world data, we instead consider candidate IVs with contrasting credibility assessments from prior studies. As shown in Table 2, the design-supported same-sex instrument yields a p-value above 0.05, providing no significant evidence against conditional exogeneity, whereas the exogeneity-challenged twin-birth instrument yields a p-value below 0.05, providing strong evidence of violation. These results are consistent with prior literature assessments of these instruments.

5.4 Sensitivity to Sample Size and Hyperparameters

We examine the sensitivity of KICS to sample size and key hyperparameters. Figure 4b shows no false detections when conditional exogeneity holds, while detection improves with increasing sample size and confounding strength. Figures 4c and 4d show similar stability across basis sizes P and regularization parameters λ: differences are mainly concentrated under weak violations, while detection becomes consistently high as confounding strengthens. Overall, KICS is stable across the examined sample sizes and hyperparameter choices, with sensitivity arising primarily in weak-violation settings, supporting practical use across reasonable choices.

6 Conclusion

We introduced the Kernelized IV Confounding Screen (KICS), a kernel-based procedure for testing the conditional exogeneity of a single candidate IV under flexible nonlinear settings without requiring auxiliary side information. By comparing ordinary and IV-weighted kernel regressions after residualizing the candidate IV, KICS yields a formal Wald-type test based on coefficient discrepancy. Across synthetic and real-data-based experiments, KICS effectively detects violations in flexible nonlinear settings while remaining competitive when the additional conditions required by existing baselines are satisfied. Future work includes extending the discrepancy-based principle to multiple candidate IVs for joint conditional-exogeneity testing and integrating KICS with representation learning or deep kernel methods to improve power in high-dimensional settings.

REFERENCES

Douglas Almond, Kenneth Y Chay, and David S Lee. The costs of low birth weight. *The Quarterly Journal of Economics*, 120(3):1031–1083, 2005.

Joshua D. Angrist and William N. Evans. Children and their parents' labor supply: Evidence from exogenous variation in family size. *American Economic Review*, 88(3):450–477, June 1998.

Joshua D. Angrist and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, NJ, 2009.

Joshua D Angrist, Guido W Imbens, and Donald B Rubin. Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434):444–455, 1996.

Benjamin F Arnold, Ayse Ercumen, Jade Benjamin-Chung, and John M Colford Jr. Brief report: negative controls to detect selection bias and measurement bias in epidemiologic studies. *Epidemiology*, 27(5):637–641, 2016.

Nachman Aronszajn. Theory of reproducing kernels. *Transactions of the American mathematical society*, 68(3):337–404, 1950.

Alexander Balke and Judea Pearl. Bounds on treatment effects from studies with imperfect compliance. *Journal of the American statistical Association*, 92(439):1171–1176, 1997.

Blai Bonet. Instrumentality tests revisited. *arXiv preprint arXiv:1301.2258*, 2013.

Zach Branson and Luke Keele. Evaluating a key instrumental variable assumption using randomization tests. *American Journal of Epidemiology*, 189(11):1412–1420, 2020.

Patrick F Burauel. Evaluating instrument validity using the principle of independent mechanisms. *Journal of Machine Learning Research*, 24(176):1–56, 2023.

Yikai Chen, Yunxin Mao, Chunyuan Zheng, Hao Zou, Shanzhi Gu, Shixuan Liu, Yang Shi, Wenjing Yang, Kun Kuang, and Haotian Wang. Detecting unobserved confounders: A kernelized regression approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 20381–20389, 2026.

Timothy G Conley, Christian B Hansen, and Peter E Rossi. Plausibly exogenous. *Review of Economics and Statistics*, 94(1):260–272, 2012.

Issa J Dahabreh and Kirsten Bibbins-Domingo. Causal inference about the effects of interventions from observational studies in medical journals. *Jama*, 331(21):1845–1853, 2024.

Oren Danieli, Daniel Nevo, Itai Walk, Bar Weinstein, and Dan Zeltzer. Negative control falsification tests for instrumental variable designs. *American Economic Review*, 116(4):1380–1414, 2026.

Helmut Farbmacher, Raphael Guber, and Johan Vikström. Increasing the credibility of the twin birth instrument. *Journal of Applied Econometrics*, 33(3):457–472, 2018.

Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *The journal of machine learning research*, 13(1):723–773, 2012.

Florian F Gunsilius. Nontestability of instrument validity under continuous treatments. *Biometrika*, 108(4):989–995, 2021.

Xichen Guo, Zheng Li, Biwei Huang, Yan Zeng, Zhi Geng, and Feng Xie. Testability of instrumental variables in additive nonlinear, non-constant effects models. *Journal of Machine Learning Research*, 27(154):1–60, 2026.

Chad Hazlett. Kernel balancing. *Statistica Sinica*, 30(3):1155–1189, 2020.

Miguel A. Hernán and James M. Robins. *Causal Inference: What If*. Chapman & Hall/CRC, Boca Raton, FL, 2020. ISBN 978-1-4200-7616-5.

Michael P Keane and Timothy Neal. A practical guide to weak instruments. *Annual Review of Economics*, 16(1):185–212, 2024.

Désiré Kédagni and Ismael Mourifié. Generalized instrumental inequalities: testing the instrumental variable independence assumption. *Biometrika*, 107(3):661–675, 2020.

Luke Keele, Qingyuan Zhao, Rachel R Kelz, and Dylan Small. Falsification tests for instrumental variable designs with an application to tendency to operate. *Medical care*, 57(2):167–171, 2019.

George Kimeldorf and Grace Wahba. Some results on tchebycheffian spline functions. *Journal of mathematical analysis and applications*, 33(1):82–95, 1971.

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

Satyam Kumar, Yelleti Vivek, Vadlamani Ravi, and Indranil Bose. A comprehensive review of causal inference in banking, finance, and insurance. *ACM Computing Surveys*, 57(12):1–36, 2025.

Jeremy Labrecque and Sonja A Swanson. Understanding the assumptions underlying instrumental variable analyses: a brief review of falsification strategies and related tools. *Current epidemiology reports*, 5(3):214–220, 2018.

Daniel J Lewis and Karel Mertens. A robust test for weak instruments for 2sls with multiple endogenous regressors. *Review of Economic Studies*, pp. rdaf103, 2025.

Marc Lipsitch, Eric Tchetgen Tchetgen, and Ted Cohen. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology*, 21(3):383–388, 2010.

Afsaneh Mastouri, Yuchen Zhu, Limor Gultchin, Anna Korba, Ricardo Silva, Matt Kusner, Arthur Gretton, and Krikamol Muandet. Proximal causal learning with kernels: Two-stage estimation and moment restriction. In *International conference on machine learning*, pp. 7512–7523. PMLR, 2021.

Ellicott C Matthay and M Maria Glymour. Causal inference challenges and new directions for epidemiologic research on the health effects of social policies. *Current Epidemiology Reports*, 9 (1):22–37, 2022.

Krikamol Muandet, Kenji Fukumizu, Bharath Sriperumbudur, and Bernhard Schölkopf. Kernel mean embedding of distributions: A review and beyond. *Foundations and Trends in Machine Learning*, 10(1–2):1–141, 2017.

Judea Pearl. *Causality*. Cambridge university press, 2009.

Judea Pearl. On the testability of causal models with latent and instrumental variables. *arXiv preprint arXiv:1302.4976*, 2013.

Bas BL Penning de Vries and Rolf HH Groenwold. Negative controls: concepts and caveats. *Statistical Methods in Medical Research*, 32(8):1576–1587, 2023.

Niklas Pfister, Peter Bühlmann, Bernhard Schölkopf, and Jonas Peters. Kernel-based tests for joint independence. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 80(1): 5–31, 2018.

Steven D Pizer. Falsification testing of instrumental variables methods for comparative effectiveness research. *Health services research*, 51(2):790–811, 2016.

Mattia Prosperi, Yi Guo, Matt Sperrin, James S Koopman, Jae S Min, Xing He, Shannan Rich, Mo Wang, Iain E Buchan, and Jiang Bian. Causal inference and counterfactual prediction in machine learning for actionable healthcare. *Nature Machine Intelligence*, 2(7):369–375, 2020.

Christoph Schultheiss, Peter Bühlmann, and Ming Yuan. Higher-order least squares: assessing partial goodness of fit of linear causal models. *Journal of the American Statistical Association*, 119(546): 1019–1031, 2024.

Xu Shi, Wang Miao, and Eric Tchetgen Tchetgen. A selective review of negative control methods in epidemiology. *Current epidemiology reports*, 7(4):190–202, 2020.

Rahul Singh, Maneesh Sahani, and Arthur Gretton. Kernel instrumental variable regression. *Advances in Neural Information Processing Systems*, 32, 2019.

James H. Stock and Motohiro Yogo. Testing for weak instruments in linear iv regression. In Donald W. K. Andrews and James H. Stock (eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, pp. 80–108. Cambridge University Press, 2005. doi: 10.1017/CBO9780511614491.006.

Zhenting Sun. Instrument validity for heterogeneous causal effects. *Journal of Econometrics*, 237(2): 105523, 2023.

Wolfgang Wiedermann and Dexin Shi. Testing the validity of instrumental variables in just-identified linear non-gaussian models. *British Journal of Mathematical and Statistical Psychology*, 79(1): 111–138, 2026.

Feng Xie, Yangbo He, Zhi Geng, Zhengming Chen, Ru Hou, and Kun Zhang. Testability of instrumental variables in linear non-gaussian acyclic causal models. *Entropy*, 24(4):512, 2022.

Qingqing Yang, Zhirong Yang, Xianming Cai, Houyu Zhao, Jinzhu Jia, and Feng Sun. Advances in methodologies of negative controls: a scoping review. *Journal of Clinical Epidemiology*, 166: 111228, 2024.

Kun Zhang, Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. Kernel-based conditional independence test and application in causal discovery. *arXiv preprint arXiv:1202.3775*, 2012.

A Theoretical Proofs

A.1 Population Coefficient Equivalence under Conditional Exogeneity

A.1.1 Coefficient discrepancy identity

Let

$$b(V) := (\phi_1(V), \dots, \phi_P(V))^\top \in \mathbb{R}^P, \qquad R_Z := Z - \mathbb{E}[Z \mid X], \qquad w := R_Z^2.$$

Define

$$A := \mathbb{E}[b(V)b(V)^\top], \qquad B := \mathbb{E}[w\, b(V)b(V)^\top].$$

Under Condition A1 stated below, A and B are invertible.

The population KLS and IVW-KLS coefficients are

$$\boldsymbol{\alpha}_V^{\mathrm{KLS}} = \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^P} \mathbb{E}\Big[\big(Y - \boldsymbol{\alpha}^\top b(V)\big)^2\Big],$$

and

$$\boldsymbol{\alpha}_V^{\mathrm{IVW}} = \arg\min_{\boldsymbol{\alpha}\in\mathbb{R}^P} \mathbb{E}\Big[\big(Y - \boldsymbol{\alpha}^\top b(V)\big)^2 w\Big].$$

Let

$$\boldsymbol{\alpha}_0 := \boldsymbol{\alpha}_V^{\mathrm{KLS}}, \qquad e_0 := Y - \boldsymbol{\alpha}_0^\top b(V).$$

By the first-order condition for KLS,

$$\mathbb{E}[b(V)e_0] = 0.$$

By the first-order condition for IVW-KLS,

$$B\boldsymbol{\alpha}_V^{\mathrm{IVW}} = \mathbb{E}[w\, b(V)Y].$$

Substituting $Y = \boldsymbol{\alpha}_0^\top b(V) + e_0$ yields

$$\mathbb{E}[w\, b(V)Y] = \mathbb{E}[w\, b(V)b(V)^\top]\boldsymbol{\alpha}_0 + \mathbb{E}[w\, b(V)e_0] = B\boldsymbol{\alpha}_0 + \mathbb{E}[w\, b(V)e_0].$$

Hence,

$$\boldsymbol{\alpha}_V^{\mathrm{IVW}} = \boldsymbol{\alpha}_V^{\mathrm{KLS}} + B^{-1}\mathbb{E}[w\, b(V)e_0].$$

Therefore,

$$\boldsymbol{\alpha}_V^{\mathrm{IVW}} = \boldsymbol{\alpha}_V^{\mathrm{KLS}} \quad \Longleftrightarrow \quad \mathbb{E}[w\, b(V)e_0] = 0.$$

A.1.2 SUFFICIENT CONDITIONS FOR COEFFICIENT EQUIVALENCE

We now state sufficient conditions under which the weighted orthogonality holds under the null hypothesis and the weighted moment is nonzero under detectable alternatives. The following conditions are introduced solely as sufficient conditions for the proof of Theorem 1; they are not part of the main-text problem formulation.

Condition A1 (Nondegeneracy of the coefficient systems). Assume that

$$A = \mathbb{E}[b(V)b(V)^\top], \qquad B = \mathbb{E}[R_Z^2 b(V)b(V)^\top]$$

are positive definite.

Condition A2 (Model-class specification in the chosen subspace). Let

$$g_\star(V) := \mathbb{E}[Y \mid V].$$

Assume that $g_\star \in \mathcal{S}_P$, where $\mathcal{S}_P$ is the chosen P-dimensional RKHS subspace spanned by $\{\phi_i\}_{i=1}^P$. Then there exists $\boldsymbol{\alpha}_\star \in \mathbb{R}^P$ such that

$$g_\star(V) = \mathbb{E}[Y \mid V] = \boldsymbol{\alpha}_\star^\top b(V).$$

Equivalently, defining

$$\varepsilon := Y - \boldsymbol{\alpha}_\star^\top b(V),$$

we have

$$\mathbb{E}[\varepsilon \mid V] = 0.$$

Condition A3 (Auxiliary outcome decomposition and no latent-weight interaction given V). Suppose, for the purpose of stating sufficient proof conditions, that the outcome admits the decomposition

$$Y = g_0(V) + h(U_Y) + \nu,$$

where $g_0(V)$ is a measurable function of V and ν is a noise term. Under $H_0 : Z \perp U_Y \mid X$, assume that the IV-induced weight does not interact with the outcome-side latent component once V is fixed, in the sense that

$$\mathbb{E}[R_Z^2 h(U_Y) \mid V] = \mathbb{E}[R_Z^2 \mid V]\,\mathbb{E}[h(U_Y) \mid V].$$

Condition A4 (Noise orthogonality). Assume that the noise term satisfies

$$\mathbb{E}[\nu \mid V, U_Y, R_Z] = 0.$$

In particular, this implies

$$\mathbb{E}[\nu \mid V] = 0 \qquad \text{and} \qquad \mathbb{E}[R_Z^2 \nu \mid V] = 0.$$

Condition A5 (Detectability under the alternative). Under the alternative $H_1 : Z \not\perp U_Y \mid X$, the violation is assumed to be detectable by the residual-IV weighting in the chosen RKHS subspace. Specifically, let

$$d(V) := \mathbb{E}[R_Z^2 e_0 \mid V], \qquad e_0 := Y - \left(\boldsymbol{\alpha}_V^{\mathrm{KLS}}\right)^\top b(V).$$

We assume that $d(V)$ has a nonzero projection onto the chosen basis, namely

$$\mathbb{E}[b(V)d(V)] \neq 0.$$

Equivalently,

$$\mathbb{E}[R_Z^2 b(V)e_0] \neq 0.$$

This condition rules out alternatives that violate conditional exogeneity but are invisible to the particular residual-IV weighting and finite-dimensional RKHS subspace used by the test.

A.1.3 PROOF OF THEOREM 1

Under Conditions A1 and A2, the population least-squares coefficient equals $\boldsymbol{\alpha}_\star$, so $\boldsymbol{\alpha}_0 = \boldsymbol{\alpha}_\star$ and

$$\boldsymbol{\alpha}_0^\top b(V) = g_\star(V) = \mathbb{E}[Y \mid V].$$

Under Condition A3,

$$Y = g_0(V) + h(U_Y) + \nu.$$

By Condition A4,

$$\mathbb{E}[\nu \mid V] = 0.$$

Taking conditional expectations given V therefore yields

$$g_\star(V) = \mathbb{E}[Y \mid V] = g_0(V) + \mathbb{E}[h(U_Y) \mid V].$$

Hence the population KLS residual can be written as

$$\begin{aligned} e_0 &= Y - \boldsymbol{\alpha}_0^\top b(V) \\ &= Y - g_\star(V) \\ &= h(U_Y) - \mathbb{E}[h(U_Y) \mid V] + \nu. \end{aligned}$$

It follows that

$$\begin{aligned} \mathbb{E}[we_0 \mid V] = \mathbb{E}[R_Z^2 h(U_Y) \mid V] - \mathbb{E}[R_Z^2 \mid V]\,\mathbb{E}[h(U_Y) \mid V] \\ + \mathbb{E}[R_Z^2 \nu \mid V]. \end{aligned}$$

By Condition A3,

$$\mathbb{E}[R_Z^2 h(U_Y) \mid V] = \mathbb{E}[R_Z^2 \mid V]\,\mathbb{E}[h(U_Y) \mid V],$$

and by Condition A4,

$$\mathbb{E}[R_Z^2 \nu \mid V] = 0.$$

Therefore,

$$\mathbb{E}[we_0 \mid V] = 0.$$

Consequently,

$$\mathbb{E}[w\, b(V) e_0] = \mathbb{E}[b(V)\,\mathbb{E}[we_0 \mid V]] = 0.$$

Applying the identity above yields

$$\boldsymbol{\alpha}_V^{\mathrm{IVW}} = \boldsymbol{\alpha}_V^{\mathrm{KLS}}.$$

Under H_1, Condition A5 implies

$$\mathbb{E}[R_Z^2 b(V) e_0] \neq 0.$$

By the coefficient discrepancy identity,

$$\boldsymbol{\alpha}_V^{\mathrm{IVW}} - \boldsymbol{\alpha}_V^{\mathrm{KLS}} = B^{-1}\mathbb{E}[R_Z^2 b(V) e_0].$$

Since B is invertible, the nonzero weighted moment implies

$$\boldsymbol{\alpha}_V^{\mathrm{IVW}} - \boldsymbol{\alpha}_V^{\mathrm{KLS}} \neq 0.$$

Therefore,

$$\boldsymbol{\alpha}_V^{\mathrm{IVW}} \neq \boldsymbol{\alpha}_V^{\mathrm{KLS}},$$

and hence there exists at least one coordinate $i \in [P]$ such that

$$\boldsymbol{\alpha}_V^{\mathrm{IVW}}[i] \neq \boldsymbol{\alpha}_V^{\mathrm{KLS}}[i].$$

This proves the theorem stated in the main text.

A.2 Closed-form empirical KLS and IVW-KLS estimators

Let

$$K = \big[b(V_1), \ldots, b(V_N)\big] \in \mathbb{R}^{P\times N}, \qquad \boldsymbol{Y} = (Y_1, \ldots, Y_N)^\top,$$

and let

$$\widehat{\Psi}_Z = \mathrm{diag}(\hat{R}^2_{Z,1}, \ldots, \hat{R}^2_{Z,N}).$$

For KLS, the empirical objective is

$$L_N^{\mathrm{KLS}}(\boldsymbol{\alpha}) = \frac{1}{N}\|\boldsymbol{Y} - K^\top\boldsymbol{\alpha}\|_2^2 + \xi\|\boldsymbol{\alpha}\|_2^2.$$

Differentiating and setting the gradient to zero yields

$$\left(KK^\top + N\xi I_P\right)\boldsymbol{\alpha} = K\boldsymbol{Y},$$

hence

$$\widehat{\boldsymbol{\alpha}}_V^{\mathrm{KLS}} = \left(KK^\top + N\xi I_P\right)^{-1} K\boldsymbol{Y}.$$

For IVW-KLS, consider

$$L_N^{\mathrm{IVW}}(\boldsymbol{\alpha}) = \frac{1}{N}(\boldsymbol{Y} - K^\top\boldsymbol{\alpha})^\top\widehat{\Psi}_Z(\boldsymbol{Y} - K^\top\boldsymbol{\alpha}) + \xi\|\boldsymbol{\alpha}\|_2^2.$$

Differentiating and setting the gradient to zero gives

$$\left(K\widehat{\Psi}_Z K^\top + N\xi I_P\right)\boldsymbol{\alpha} = K\widehat{\Psi}_Z\boldsymbol{Y},$$

so

$$\widehat{\boldsymbol{\alpha}}_V^{\mathrm{IVW}} = \left(K\widehat{\Psi}_Z K^\top + N\xi I_P\right)^{-1} K\widehat{\Psi}_Z\boldsymbol{Y}.$$

This proves Theorem 2.

A.3 Regularity conditions, asymptotic distribution, and Wald test

Regularity conditions. We impose the following regularity conditions:

- (C1) The basis size P is fixed, and $\{(Y_i, T_i, X_i, Z_i)\}_{i=1}^N$ are i.i.d.
- (C2) $\mathbb{E}\|b(V)\|_2^8 < \infty$, $\mathbb{E}|R_Z|^8 < \infty$, and $\mathbb{E}|Y|^8 < \infty$.
- (C3) $A = \mathbb{E}[b(V)b(V)^\top]$ and $B = \mathbb{E}[R_Z^2 b(V)b(V)^\top]$ are positive definite.
- (C4) Under H_0, Conditions A1–A4 in Appendix A.1.2 hold, so that
$$\boldsymbol{\alpha}_V^{\mathrm{KLS}} = \boldsymbol{\alpha}_V^{\mathrm{IVW}} =: \boldsymbol{\alpha}_0.$$
- (C5) Let $e_i := Y_i - \boldsymbol{\alpha}_0^\top b(V_i)$, $\hat{R}_{Z,i} := Z_i - \hat{m}(X_i)$. The first-stage estimator $\hat{m}(X)$ is constructed such that
$$\left\|\frac{1}{N}\sum_{i=1}^N\left(\hat{R}^2_{Z,i} - R^2_{Z,i}\right)b(V_i)e_i\right\|_2 = o_p(N^{-1/2}),$$
and
$$\left\|\frac{1}{N}\sum_{i=1}^N\left(\hat{R}^2_{Z,i} - R^2_{Z,i}\right)b(V_i)b(V_i)^\top\right\|_{\mathrm{op}} = o_p(1).$$
These conditions can be facilitated by suitable sample splitting or cross-fitting together with sufficiently fast first-stage convergence and appropriate moment conditions.
- (C6) $\xi = \xi_N \to 0$ and $\sqrt{N}\,\xi_N \to 0$.
- (C7) The asymptotic covariance matrix of the coefficient discrepancy is positive definite.
- (C8) Let $\hat{\zeta}_i$ denote the plug-in influence contribution defined below. Assume that
$$\frac{1}{N}\sum_{i=1}^N\left\|\hat{\zeta}_i - \zeta_i\right\|_2^2 = o_p(1).$$

Proof of asymptotic normality. Under H_0 and Condition (C4), define

$$e_i := Y_i - \boldsymbol{\alpha}_0^\top b(V_i), \qquad \psi_i^{\mathrm{KLS}} := b(V_i)e_i, \qquad \psi_i^{\mathrm{IVW}} := R_{Z,i}^2 b(V_i)e_i.$$

In particular, Condition (C4) ensures that the weighted orthogonality proved in Appendix A.1.3 holds at the population level. By Condition (C6), the contribution of ridge regularization is $o_p(1)$ at the root-N scale.

Standard M-estimation expansions yield

$$\sqrt{N}\left(\widehat{\boldsymbol{\alpha}}_V^{\mathrm{KLS}} - \boldsymbol{\alpha}_0\right) = A^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^N \psi_i^{\mathrm{KLS}} + o_p(1),$$

and

$$\sqrt{N}\left(\widehat{\boldsymbol{\alpha}}_V^{\mathrm{IVW}} - \boldsymbol{\alpha}_0\right) = B^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^N \psi_i^{\mathrm{IVW}} + o_p(1).$$

By Condition (C5), replacing $R_{Z,i}$ by $\hat{R}_{Z,i}$ only contributes an $o_p(1)$ remainder after normalization by $\sqrt{N}$.

Subtracting the two expansions,

$$\sqrt{N}\left(\widehat{\boldsymbol{\alpha}}_V^{\mathrm{IVW}} - \widehat{\boldsymbol{\alpha}}_V^{\mathrm{KLS}}\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^N \zeta_i + o_p(1),$$

where

$$\zeta_i := B^{-1}\psi_i^{\mathrm{IVW}} - A^{-1}\psi_i^{\mathrm{KLS}}.$$

By Condition (C4), ζ_i are mean-zero under H_0, while Conditions (C1)–(C3) ensure that they are i.i.d. with finite covariance. Therefore, the multivariate central limit theorem implies

$$\sqrt{N}\left(\widehat{\boldsymbol{\alpha}}_V^{\mathrm{IVW}} - \widehat{\boldsymbol{\alpha}}_V^{\mathrm{KLS}}\right) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \Sigma_Z),$$

where

$$\Sigma_Z := \mathrm{Var}(\zeta_i).$$

As an immediate consequence, under H_0,

$$\left\|\widehat{\boldsymbol{\alpha}}_V^{\mathrm{IVW}} - \widehat{\boldsymbol{\alpha}}_V^{\mathrm{KLS}}\right\|_2 = O_p(N^{-1/2}).$$

Thus, the estimated coefficient discrepancy vanishes at the root-N rate under conditional exogeneity.

Wald-test details. We first construct a consistent estimator of the asymptotic covariance matrix. A natural plug-in estimator is

$$\widehat{\Sigma}_Z = \frac{1}{N}\sum_{i=1}^N \hat{\zeta}_i\hat{\zeta}_i^\top,$$

where

$$\hat{\zeta}_i = \widehat{B}^{-1}\hat{\psi}_i^{\mathrm{IVW}} - \widehat{A}^{-1}\hat{\psi}_i^{\mathrm{KLS}},$$

with

$$\widehat{A} = \frac{1}{N}\sum_{i=1}^N b(V_i)b(V_i)^\top, \qquad \widehat{B} = \frac{1}{N}\sum_{i=1}^N \hat{R}_{Z,i}^2 b(V_i)b(V_i)^\top,$$

and

$$\hat{\psi}_i^{\mathrm{KLS}} = b(V_i)\hat{e}_i, \qquad \hat{\psi}_i^{\mathrm{IVW}} = \hat{R}_{Z,i}^2 b(V_i)\hat{e}_i, \qquad \hat{e}_i = Y_i - \left(\widehat{\boldsymbol{\alpha}}_V^{\mathrm{KLS}}\right)^\top b(V_i).$$

To establish consistency, define

$$\delta_i := \hat{\zeta}_i - \zeta_i.$$

Then

$$\hat{\zeta}_i\hat{\zeta}_i^\top - \zeta_i\zeta_i^\top = \delta_i\zeta_i^\top + \zeta_i\delta_i^\top + \delta_i\delta_i^\top.$$

Therefore,

$$\left\| \widehat{\Sigma}_Z - \frac{1}{N}\sum_{i=1}^{N} \zeta_i \zeta_i^\top \right\|_{\mathrm{op}} \le 2 \left(\frac{1}{N}\sum_{i=1}^{N} \|\delta_i\|_2^2 \right)^{1/2} \left(\frac{1}{N}\sum_{i=1}^{N} \|\zeta_i\|_2^2 \right)^{1/2} + \frac{1}{N}\sum_{i=1}^{N} \|\delta_i\|_2^2.$$

By Condition (C8),

$$\frac{1}{N}\sum_{i=1}^{N} \|\delta_i\|_2^2 = o_p(1).$$

Under the moment conditions,

$$\frac{1}{N}\sum_{i=1}^{N} \|\zeta_i\|_2^2 = O_p(1).$$

Hence

$$\widehat{\Sigma}_Z - \frac{1}{N}\sum_{i=1}^{N} \zeta_i \zeta_i^\top = o_p(1).$$

Moreover, under H_0, $\mathbb{E}[\zeta_i] = 0$, and the law of large numbers gives

$$\frac{1}{N}\sum_{i=1}^{N} \zeta_i \zeta_i^\top \xrightarrow{p} \mathbb{E}[\zeta_i \zeta_i^\top] = \Sigma_Z.$$

Therefore,

$$\widehat{\Sigma}_Z \xrightarrow{p} \Sigma_Z.$$

We then form the Wald statistic

$$Q_N = N \left(\widehat{\boldsymbol{\alpha}}_V^{\mathrm{IVW}} - \widehat{\boldsymbol{\alpha}}_V^{\mathrm{KLS}} \right)^\top \widehat{\Sigma}_Z^{-1} \left(\widehat{\boldsymbol{\alpha}}_V^{\mathrm{IVW}} - \widehat{\boldsymbol{\alpha}}_V^{\mathrm{KLS}} \right).$$

Under Condition (C7), $\widehat{\Sigma}_Z$ is nonsingular with probability approaching one. Combining the asymptotic normality result with $\widehat{\Sigma}_Z \xrightarrow{p} \Sigma_Z$ yields, under H_0,

$$Q_N \xrightarrow{D} \chi_P^2.$$

Hence KICS rejects H_0 whenever

$$Q_N > \chi_{P,1-\alpha}^2.$$

This proves Theorem 3.

A.4 Complexity derivation

We now justify the complexity bound in the main text. Residualizing the candidate IV incurs cost $C_m(N, d_X)$, depending on the chosen first-stage regression model. To construct the kernel design matrix, KICS evaluates the kernel between the P selected basis points and all N observations in $V = (T, X)$, directly forming $K \in \mathbb{R}^{P \times N}$. When each kernel evaluation scales linearly with the input dimension d_V, this step costs $O(PNd_V)$.

Forming $KK^\top$ and $K\widehat{\Psi}_Z K^\top$ each costs $O(P^2 N)$, so coefficient estimation contributes $O(P^2 N)$ for these matrix products. Solving the two $P \times P$ regularized linear systems costs $O(P^3)$. The covariance estimation and Wald statistic require at most $O(P^2 N + P^3)$ additional operations and therefore do not change the overall order.

Therefore, the total complexity of KICS is

$$O\big(C_m(N, d_X) + PNd_V + P^2 N + P^3\big),$$

which proves the claim in the main text.

B Additional Experimental Details

B.1 Experimental Settings and Data Details

This section provides further details on the synthetic regimes, the Twins semi-synthetic benchmark, and the Angrist–Evans real-world application. Unless otherwise stated, synthetic and semi-synthetic experiments use sample size $N = 300$. Throughout, let $c \geq 0$ denote the latent confounding strength and let

$$\sigma(a) = \frac{1}{1 + \exp(-a)}$$

denote the logistic function. We use Z below for the candidate IV.

General nonlinear setting. We generate

$$X_1, X_2 \overset{\text{i.i.d.}}{\sim} \text{Unif}(-1, 1), \qquad U = c\varepsilon_U, \qquad \varepsilon_U \sim \mathcal{N}(0, 1),$$

and define

$$s(X) = 0.6X_1 - 0.5X_2 + 0.4X_1X_2.$$

A binary IV precursor is generated as

$$Z_b \sim \text{Bernoulli}(\sigma(s(X) + 0.8U)),$$

from which the continuous candidate IV is constructed as

$$Z = Z_b + 0.05\varepsilon_Z.$$

The treatment and outcome follow

$$T = 1.2Z + 0.6X_1^2 + 0.4\sin(\pi X_2) + 0.8U + 0.25\varepsilon_T,$$

and

$$Y = 0.8T + 0.6T^2 + 0.5\sin(\pi X_1) + 0.4X_2^2 + U + 0.35\varepsilon_Y,$$

where $\varepsilon_Z, \varepsilon_T, \varepsilon_Y$ are independent standard Gaussian noises. At $c = 0$, the latent component vanishes and conditional exogeneity holds; increasing c induces progressively stronger dependence between the candidate IV and outcome-side latent confounding. This nonlinear continuous DGP is used for both the general and auxiliary-information regimes.

Auxiliary-information regimes. For methods requiring auxiliary side information, the nonlinear DGP above additionally generates method-specific auxiliary quantities. For the matched regime, the assignment probability is

$$\pi(X) = \sigma(s(X)),$$

and an observed balance proxy is generated as

$$B = 0.8U + 0.5\varepsilon_B.$$

The Branson-style test is supplied with the augmented balance variables

$$(X_1, X_2, B, X_1X_2).$$

For the negative-control test, we generate

$$C = 0.8\cos(X_1) + 0.4X_2^2 + U + 0.25\varepsilon_C,$$

while the AIT auxiliary variable is constructed as

$$A = Y - \left[0.8T + 0.6T^2 + 0.5\sin(\pi X_1) + 0.4X_2^2\right].$$

The general mismatch regime uses the same primary DGP but does not provide these valid auxiliary ingredients. Instead, the Branson-style test receives only (X_1, X_2) without the latent balance proxy, the negative-control variable is replaced by

$$C_{\text{mis}} = 0.8\cos(X_1) + 0.4X_2^2 + 0.25\varepsilon_C,$$

and AIT uses

$$A_{\text{mis}} = 0.5\cos(X_1) + 0.2X_2 + 0.25\varepsilon_A.$$

This construction allows the same underlying nonlinear system to be evaluated both when valid auxiliary information is available and when the ingredients required by auxiliary-information baselines are unavailable or misspecified.

Linear non-Gaussian setting. To construct a regime favorable to linear non-Gaussian baselines, we generate

$$X = (X_1, X_2)^\top \sim \mathcal{N}(0, I_2), \qquad U = c\varepsilon_U, \qquad \varepsilon_U \sim \text{Laplace}(0, 1),$$

together with independent noise terms

$$\varepsilon_Z, \varepsilon_T, \varepsilon_Y \overset{\text{i.i.d.}}{\sim} \text{Laplace}(0, 0.4).$$

The candidate IV, treatment, and outcome are generated as

$$Z = 0.8X_1 - 0.6X_2 + U + \varepsilon_Z,$$

$$T = 1.3Z + 0.5X_1 + 0.8U + \varepsilon_T,$$

and

$$Y = 1.2T + 0.5X_2 + U + \varepsilon_Y.$$

Thus, $c = 0$ yields a conditionally exogenous candidate IV, whereas increasing c introduces shared latent variation between Z and the outcome. The resulting system satisfies the linear and non-Gaussian structure targeted by the corresponding specialized baselines.

Discrete-support setting. For the discrete regime, we generate

$$X_1, X_2 \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(0.5),$$

and use a bounded asymmetric latent factor

$$U = B_U - 0.35, \qquad B_U \sim \text{Bernoulli}(0.35),$$

so that $\mathbb{E}[U] = 0$. The candidate IV is generated according to

$$Z \sim \text{Bernoulli}(p_Z),$$

where

$$p_Z = \sigma(-0.7 + 0.8X_1 - 0.6X_2 + 0.8X_1X_2 + 2cU)\,.$$

The binary treatment follows

$$T \sim \text{Bernoulli}(p_T),$$

with

$$p_T = \sigma(-0.2 + 1.3Z + 0.5X_1 - 0.3X_2 + 0.6X_1X_2 + 0.7U)\,,$$

and the binary outcome is generated as

$$Y \sim \text{Bernoulli}(p_Y),$$

where

$$p_Y = \sigma(-0.5 + 1.2T + 0.7X_1X_2 + 2cU)\,.$$

At $c = 0$, the latent factor does not directly enter either the candidate-IV or outcome equation, so the conditional-exogeneity null is preserved. For $c > 0$, the shared $2cU$ terms induce dependence between the candidate IV and outcome-side latent confounding while retaining binary support for Z, T, and Y. This regime therefore matches the discrete-support conditions required by the corresponding specialized baselines.

Twins semi-synthetic setting. For the Twins benchmark (Almond et al., 2005), we retain observed covariates from the original data and generate the candidate IV, treatment, and outcome semi-synthetically. Specifically, we use standardized maternal age, gestational age, prenatal-visit information, and adequacy of care as

$$X = (X_1, X_2, X_3, X_4).$$

We further construct standardized mean birth weight M and within-pair birth-weight difference D from the observed twin-pair records. For each repetition, observations are sampled without replacement and

$$U \sim \mathcal{N}(0, 1), \qquad g(X) = 0.70X_1X_2 + 0.15X_3^2.$$

The candidate IV is generated as

$$Z = 0.95D + 0.15X_1 - 0.08X_2 + 0.85cU + 0.06\varepsilon_Z,$$

the treatment as

$$T = 1.10Z + 0.25M + 0.15g(X) + 0.06\varepsilon_T,$$

and the outcome as

$$Y = 1.15T + 0.18g(X) + 2.40cU + 0.06\varepsilon_Y,$$

where $\varepsilon_Z, \varepsilon_T, \varepsilon_Y$ are independent standard Gaussian noises. At $c = 0$, the candidate IV contains no latent component shared with the outcome, whereas $c > 0$ introduces a common latent channel through U. This construction preserves the empirical covariate and birth-weight structure of Twins while allowing the strength of conditional-exogeneity violations to be controlled explicitly.

Angrist–Evans real-world setting. For the real-world candidate-IV screening experiment, we use the 1980 U.S. Census application from Angrist and Evans (Angrist & Evans, 1998). The analysis follows the fertility-IV setting, where the treatment is whether the mother has more than two children and the outcome is maternal labor supply measured by weeks worked. Observed covariates include maternal demographic characteristics and fertility-related variables.

We evaluate two candidate IVs with contrasting credibility assessments in the literature. The first is the same-sex composition instrument, defined by whether the first two children have the same sex. This instrument is design-supported because sibling sex composition provides an approximately random source of fertility variation (Angrist & Evans, 1998). The second is the second-parity twin-birth instrument, which generates variation in family size but has documented exogeneity concerns due to associations between twinning and outcome-related characteristics (Farbmacher et al., 2018).

These two literature-based candidate IVs provide a real-world setting to evaluate KICS for screening IV conditional exogeneity.

KICS implementation. For residualization, we estimate $\mathbb{E}[Z \mid X]$ using ridge regression with cross-fitted polynomial features. We use two-fold cross-fitting and ridge parameter 10^{-3}, with quadratic features in the nonlinear and discrete synthetic settings and linear features in the linear, Twins, and Angrist–Evans settings.

B.2 Baseline details

B.2.1 Baselines requiring structural restrictions

Linear-nonGaussian-IV (LinNG-IV). LinNG-IV is based on the just-identified linear non-Gaussian IV validity test (Wiedermann & Shi, 2026). It exploits properties of linear non-Gaussian models to test whether latent confounding between a candidate IV and the outcome is present. Its intended regime is linear and non-Gaussian, and its main use in our experiments is to represent methods that become available only after strong structural restrictions are imposed.

IV-PIM. IV-PIM is based on the Principle of Independent Mechanisms perspective for evaluating instrument validity (Burauel, 2023). It assesses whether the statistical dependencies observed in the data are compatible with a valid IV under a linear structural regime. In our comparisons, it represents methods whose validity checks rely on linear structure rather than flexible nonlinear modeling.

IV-GIN. IV-GIN is based on the IV-GIN condition in linear non-Gaussian acyclic causal models (Xie et al., 2022). It uses the generalized independent noise criterion as a necessary condition for IV validity. In our experiments, it serves as a representative of methods that require both linear structure and non-Gaussianity to obtain testable implications.

Kitagawa. Kitagawa's test evaluates instrument validity in heterogeneous treatment effect models (Kitagawa, 2015). It is derived from testable implications that arise when the treatment is binary and the instrument is discrete. In our experimental setting, continuous variables are discretized when necessary so that this baseline can be applied.

Sun. Sun's nonparametric validity test targets heterogeneous causal effects with multivalued ordered or unordered treatments (Sun, 2023). It generalizes discrete-support validity testing beyond the binary-treatment setting. As with Kitagawa, continuous variables are discretized when necessary to place the problem in the regime for which the baseline is designed.

B.2.2 BASELINES REQUIRING AUXILIARY SIDE INFORMATION

Branson-style randomization test (Branson). Branson follows the randomization-test perspective for IV falsification (Branson & Keele, 2020). It evaluates whether the candidate IV behaves as if randomly assigned, conditional on observed covariates, under a user-specified assignment mechanism. Its applicability therefore depends on whether such a mechanism can be credibly specified.

Negative-control falsification test (NCF). NCF uses negative controls as proxies for threats to IV validity (Danieli et al., 2026). It tests whether the candidate IV exhibits conditional independence patterns that should hold if the negative controls successfully capture relevant threats. Its applicability depends on the availability and credibility of suitable negative-control variables.

Auxiliary-based Independence Test (AIT). AIT uses an auxiliary construction to derive a testable implication for IV validity in additive nonlinear, non-constant-effects models (Guo et al., 2026). In our comparisons, it represents methods that avoid linearity but still require additional auxiliary structure beyond the observed quadruple (Y, T, X, Z).

C ADDITIONAL EXPERIMENTS

C.1 RUNTIME COMPARISON

We additionally report the computational runtime of KICS and the competing methods in the general nonlinear experiment corresponding to Figure 3a. Runtime is measured per method call with $N = 300$, excluding data-generation time, and averaged over 180 runs under the same computational environment. As shown in Table 3, KICS requires approximately 0.003 seconds per test, remaining in the same millisecond-scale range as most lightweight baselines while being substantially faster than Branson, Kitagawa, and Sun. These results indicate that the broader applicability of KICS does not incur substantial computational overhead in this setting.

Table 3: Average runtime per method call in the general nonlinear setting.

Method	**Runtime (s)**
KICS (Ours)	0.0030
LinNG-IV	0.0014
IV-GIN	0.0021
IV-PIM	0.0016
Kitagawa	0.4888
Sun	0.4567
Branson	0.0129
NCF	0.0012
AIT	0.0011

D AI ASSISTANCE STATEMENT

AI-based tools were used solely to assist with language polishing and presentation. All AI-assisted text was reviewed and revised by the authors, who take full responsibility for the final manuscript.