Econometric Theory, **0**, 2026, 1–25.
doi:10.1017/S0266466626100437

A CONSISTENT ICM-BASED χ^2 SPECIFICATION TEST

FEIYU JIANG
Fudan University

EMMANUEL SELORM TSYAWO
University of Alabama

In spite of the omnibus property of integrated conditional moment (ICM) specification tests, they are not commonly used in empirical practice owing to features such as the non-pivotality of the test and the high computational cost of available bootstrap schemes, especially in large samples. This article proposes specification and mean independence tests based on ICM metrics. The proposed test exhibits consistency, asymptotic χ^2-distribution under the null hypothesis, and computational efficiency. Moreover, it demonstrates robustness to heteroskedasticity of unknown form and can be adapted to enhance power toward specific alternatives. A power comparison with classical bootstrap-based ICM tests using Bahadur slopes is also provided. Monte Carlo simulations are conducted to showcase the excellent size control and competitive power of the proposed test.

1. INTRODUCTION

Model misspecification is a major source of misleading inference in empirical work. This issue is further compounded when various competing models are available. It is thus imperative that model-based statistical inference be accompanied by proper model checks such as specification tests (Stute, 1997).

Existing tests in the specification testing literature can be categorized into three classes, namely, conditional moment (CM) tests, nonparametric tests, and integrated CM (ICM) tests. The class of CM tests, such as those proposed by Newey (1985) and Tauchen (1985), is not consistent as it relies on only a finite number of moment conditions implied by the null hypothesis (Bierens, 1990). The class of nonparametric tests is therefore proposed as a remedy (see, e.g., Wooldridge, 1992; Yatchew, 1992; Hardle and Mammen, 1993; Hong and White, 1995; Zheng, 1996; Li and Wang, 1998; Fan and Li, 2000; Su and White, 2007; Li, Liao, and Zhou, 2026). The key idea of these test statistics is to nonparametrically estimate the CMs—such as through local smoothing techniques—and then compare them to their parametric counterparts under the null hypothesis.

The authors thank the Editor, Prof. Peter C. B. Phillips, the Co-Editor, Prof. Liangjun Su, and three anonymous referees for their constructive comments, which greatly improved the quality of this work. Address correspondence to Emmanuel Selorm Tsyawo, Department of Economics, Finance and Legal Studies, Culverhouse College of Business, University of Alabama, United States, e-mail: estsyawo@gmail.com.

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0266466626100437 Published online by Cambridge University Press

The class of nonparametric tests may encounter challenges, such as nonparametric smoothing and suboptimal performance stemming from over-fitting the nonparametric alternative. In contrast, the class of ICM tests, such as those introduced by Bierens (1982, 1990), Delgado (1993), Bierens and Ploberger (1997), Stute (1997), Delgado, Domínguez, and Lavergne (2006), Escanciano (2006a), Domínguez and Lobato (2015), Su and Zheng (2017), and Antoine and Lavergne (2023), has gained popularity due to its ability to avoid these issues and detect local alternatives at faster rates. ICM metrics, on which ICM tests are based, also appear in other contexts: martingale difference hypothesis tests (Escanciano, 2009a), joint coefficient and specification tests (Antoine and Lavergne, 2023), model-free feature screening (Zhu et al., 2011; Shao and Zhang, 2014; Li et al., 2023), model estimation (Escanciano, 2018; Tsyawo, 2023), specification tests of the propensity score (Sant'Anna and Song, 2019), and tests of the instrumental variable (IV) relevance condition in ICM estimators (Escanciano, 2018; Tsyawo, 2023).

Despite their advantages, ICM tests are not widely used in empirical research (Escanciano, 2009b; Domínguez and Lobato, 2015). First, ICM test statistics are not pivotal under the null hypothesis, thus critical values cannot be tabulated analytically (Bierens and Ploberger, 1997; Domínguez and Lobato, 2015). Second, ICM tests—when implemented via the wild bootstrap—tend to be computationally costly, as they require estimations on resampled data to compute p-values. Moreover, the wild bootstrap-based ICM test is arguably unsuitable for limited dependent outcome variable models such as logit, because bootstrap replicates of the outcome may fail to respect the outcome variable's limited support. While the more recent multiplier bootstrap (MB) approach to ICM specification testing (e.g., Escanciano, 2009b, 2024; Li and Song, 2022) offers a substantial computational advantage relative to the wild bootstrap by avoiding model re-estimation, it still entails non-negligible computational complexity due to re-sampling and the removal of the effect of estimating nuisance parameters. Third, although ICM tests are omnibus (Bierens, 1982; Stute, 1997; Domínguez and Lobato, 2015), they only have substantial local power against alternatives in a finite-dimensional space (Escanciano, 2009a). Moreover, it is not obvious how to leverage prior knowledge of potential directions under the alternative to enhance the power of existing ICM tests.

This article proposes a consistent χ^2-test for the unified ICM framework of mean independence and specification testing. The key idea is to augment the ICM metric with a user-specified non-degenerate transformation of the conditioning covariates as in CM tests, which removes the first-order degeneracy inherent in classical ICM tests and yields a pivotal test statistic. The approach accommodates endogenous regressors, IVs, and heteroskedasticity of unknown form in both linear and non-linear models. As an ICM-based test, it is omnibus and capable of detecting a wide range of model misspecifications, including violations of IV exogeneity. Similar to the MB method (Escanciano, 2009b), our approach is

https://doi.org/10.1017/S0266466626100437 Published online by Cambridge University Press

suited for models with possibly non-additively separable errors or non-continuous outcomes.

Compared to existing ICM tests, our test has three advantages. First, it can be implemented as a χ^2- or two-sided t-test, which can be interpreted more easily when compared to bootstrap-based tests. Second, the proposed test does not require bootstrap calibration of critical values; hence, it is computationally fast and remains feasible even in very large samples. Third, although our proposed test is not optimal, its power can be enhanced with the knowledge of directions under the alternative or, more generally, with directions the researcher may have in mind, whereas ICM tests lack this property. Therefore, we consider our test as a bridge between CM tests and ICM tests. We also acknowledge that the implementation of our test involves a tuning parameter, which is used in computing the generalized inverse that forms the Wald-type test statistic.

This article is not the first attempt at circumventing the non-pivotality of ICM test statistics. Bierens (1982) approximates the critical values of the ICM specification test using Chebyshev's inequality for first moments under the null hypothesis, which is subsequently improved upon by Bierens and Ploberger (1997). Bierens (1982) also proposes a χ^2-test based on two estimates of Fourier coefficients and a carefully chosen tuning parameter. Simulation evidence therein shows a high level of sensitivity to the tuning parameter. Besides, estimating Fourier coefficients no longer makes the test ICM as the test statistic is no longer "integrated." Another attempt in the literature is the conditional Monte Carlo approach of De Jong (1996) and Hansen (1996). These are, however, computationally costly as Bierens and Ploberger (1997) notes. Recent studies have also revisited the goal of constructing pivotal and computationally feasible specification tests within the CM framework. Raiola (2024) draws inspiration from the classical Pearson's χ^2 test by partitioning the sample space into cells. Although this test enjoys pivotality and excellent computational scalability, its power is inherently tied to the chosen partition and is therefore not fully omnibus: smooth model deviations that do not necessarily manifest as changes in constructed cells may go undetected. Li and Song (2025) leverage sample splitting and learn the optimal projection direction in the reproducing kernel Hilbert space via support vector machines (SVMs). This means that it can achieve high local power when the learned projection aligns with the alternative; however, sample splitting and hyperparameter tuning (in both kernels and SVM) may introduce finite-sample variability.

The rest of the article is organized as follows. Section 2 provides a brief presentation of ICM metrics, a new metric, and its omnibus property. Section 3 proposes the χ^2-test statistic and derives its limiting distribution under the null, local, and fixed alternative hypotheses within the unified framework of mean independence and specification tests. Monte Carlo simulations in Section 4 compare the empirical size and power of the χ^2 specification test to the multiplier and wild bootstrap-based ICM specification tests, and Section 5 concludes. All technical proofs and additional simulation results are relegated to the Supplementary Material.

https://doi.org/10.1017/S0266466626100437 Published online by Cambridge University Press

Notation: For $a \in \mathbb{R}^p$, we denote its transpose by $a^\top$, and its Euclidean norm as $\|a\|$. We denote i as the imaginary unit, which satisfies $\mathrm{i}^2 = -1$. "$\xrightarrow{p}$" and "$\xrightarrow{d}$" denote convergence in probability and distribution, respectively. Throughout the article, for a random vector W, we denote $W^\dagger$ as its independent and identically distributed (*iid*) copy, and write $\mathbb{E}_n W = n^{-1}\sum_{i=1}^n W_i$ as the empirical mean for *iid* copies $\{W_i\}_{i=1}^n$ of W. To cut down on notational clutter, $\widetilde{W}$ is sometimes used to denote the centered version of a random variable W, that is, $\widetilde{W} := W - \mathbb{E}W$.

2. THE FRAMEWORK

In this section, we briefly discuss ICM metrics, their omnibus property, and a new omnibus metric that characterizes mean independence.

2.1. ICM Metrics

For a random variable $U \in \mathbb{R}$ and a random vector $Z \in \mathbb{R}^{p_z}$, we say that U is mean-independent of Z, if

$$\mathbb{E}[U \mid Z] = \mathbb{E}[U] \textit{ almost surely } (a.s.), \tag{1}$$

otherwise, U is mean-dependent on Z. To characterize the relationship (1), the existing literature puts much effort into studying ICM mean dependence metrics of the form:

$$T(U \mid Z; \nu) = \int_\Pi \left|\mathbb{E}[(U - \mathbb{E}U)w(s,Z)]\right|^2 \nu(ds),$$

where $w(s,Z)$ is a weight function indexed by $s \in \Pi$, and ν is a measure on Π. A notable feature of $T(U \mid Z; \nu)$ is its *omnibus property*, namely, $T(U \mid Z; \nu) = 0$ if and only if (1) holds (see, e.g., Shao and Zhang, 2014, Thm. 1.2). The omnibus property guarantees the consistency of ICM tests. Therefore, a larger value of $T(U \mid Z; \nu)$ indicates a stronger mean dependence of U on Z.

Under suitable conditions, the ICM metric has the general form:

$$\begin{aligned} T(U \mid Z; \nu) &= \int_\Pi \Big(\mathbb{E}[(U - \mathbb{E}U)w(s,Z)]\Big)\overline{\Big(\mathbb{E}[(U - \mathbb{E}U)w(s,Z)]\Big)}\nu(ds) \\ &= \mathbb{E}\Big[(U - \mathbb{E}U)(U^\dagger - \mathbb{E}U)\int_\Pi \big(w(s,Z)\overline{w(s,Z^\dagger)}\big)\nu(ds)\Big] \\ &:= \mathbb{E}\Big[(U - \mathbb{E}U)(U^\dagger - \mathbb{E}U)K(Z,Z^\dagger)\Big] \\ &:= \mathrm{ICM}(U \mid Z), \end{aligned}$$

where $\overline{w(\cdot,\cdot)}$ denotes the complex conjugate of $w(\cdot,\cdot)$ and $K(\cdot,\cdot)$ denotes the kernel. Table 1 provides a few examples of commonly used ICM kernels (see Li et al., 2023 for a recent review).

https://doi.org/10.1017/S0266466626100437 Published online by Cambridge University Press

TABLE 1. Examples of ICM kernels

$w(s,Z)$	$\nu(ds)$	$K(Z,Z^\dagger)$	Reference
$\exp(\mathrm{i}Z^\top s)$	$\prod_{l=1}^{p_Z}\phi(s_l)ds_l$	$\exp\left(-0.5\|Z-Z^\dagger\|^2\right)$	Bierens (1982)
$1(Z\le s)$	$dF_Z(s)$	$1-F_Z(Z\vee Z^\dagger)$	Stute (1997)
$\exp(\mathrm{i}Z^\top s)$	$\left(c_{p_Z}\|\|s\|\|^{1+p_Z}\right)^{-1}ds$	$-\|Z-Z^\dagger\|$	Shao and Zhang (2014)

Note: $\phi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$ is the standard normal probability density function, and $c_p=\frac{\pi^{(1+p)/2}}{\Gamma((1+p)/2)}$, $p\ge 1$, where $\Gamma(\cdot)$ is the complete gamma function.

Examples of weight functions from the literature include the step function $w(s,Z)=\mathrm{I}(Z\le s)$ (e.g., Stute, 1997; Domínguez and Lobato, 2004; Delgado et al., 2006; Escanciano, 2006b; Zhu et al., 2011); a one-dimensional projection in the step function $w(s,Z)=\mathrm{I}(Z^\top s_{-1}\le s_1)$ (e.g., Escanciano, 2006a; Kim, Balakrishnan, and Wasserman, 2020); the real exponential $w(s,Z)=\exp(Z^\top s)$ (e.g., Bierens, 1990); and the complex exponential $w(s,Z)=\exp(\mathrm{i}Z^\top s)$ (e.g., Bierens (1982), Shao and Zhang (2014), and Antoine and Lavergne (2023)). The space Π in Escanciano (2006a) and Kim et al. (2020) is given by $\Pi=\mathbb{R}\times\mathbb{S}_{p_z}$, where $\mathbb{S}_{p_z}$ denotes the space of $p_z\times 1$ vectors with unit Euclidean norm while $\Pi=\mathbb{R}^{p_z}$ for the other works mentioned above. See Escanciano (2006b, Lemma 1) for a general characterization of ICM weight functions.

The omnibus property of ICM metrics thus translates as

$$\mathrm{ICM}(U\mid Z)=0 \quad \text{if and only if} \quad \mathbb{E}[U\mid Z]=\mathbb{E}[U] \text{ } \textit{almost surely} \text{ } (a.s). \tag{2}$$

When U and Z are observed, a natural empirical estimator of $\mathrm{ICM}(U\mid Z)$ is given by the (modified) U-statistic

$$\widehat{\mathrm{ICM}}_n(U\mid Z)=\frac{1}{n(n-1)}\sum_{i\ne j}(U_i-\mathbb{E}_n[U])(U_j-\mathbb{E}_n[U])K(Z_i,Z_j).$$

As with most existing ICM-based tests, the asymptotic null distribution of the corresponding test statistic (under (1)) is non-pivotal (Bierens and Ploberger, 1997, Thm. 3):

$$n\widehat{\mathrm{ICM}}_n(U\mid Z)\xrightarrow{d}\sum_{k=1}^{\infty}\lambda_k G_k^2, \tag{3}$$

granted $\mathbb{E}[U^2U^{\dagger 2}K^2(Z,Z^\dagger)]<\infty$. Here, $\{G_k\}_{k=1}^\infty$ is a sequence of *iid* standard Gaussian random variables, and $\{\lambda_k\}_{k=1}^\infty$, $\lambda_k\ge 0$, is a sequence of non-increasing coefficients that depend on the distribution of $[U,Z^\top]^\top$ and the kernel $K(\cdot,\cdot)$. Thus, the limiting distribution of $n\widehat{\mathrm{ICM}}_n(U\mid Z)$ under (1) is non-pivotal, and re-sampling/bootstrap techniques are usually required to obtain valid p-values for inference.

https://doi.org/10.1017/S0266466626100437 Published online by Cambridge University Press

2.2. A New Characterization of Mean Independence

Consider the following test hypotheses of mean independence:

$$\begin{aligned}&\mathbb{H}_o : \mathbb{E}[U \mid Z] - \mathbb{E}[U] = 0 \ a.s.;\\ &\mathbb{H}_a : \mathbb{P}\big(\mathbb{E}[U \mid Z] = \mathbb{E}[U]\big) < 1.\end{aligned} \tag{4}$$

The above hypotheses of interest, in view of the omnibus property (2), can be restated as testing $\mathrm{ICM}(U \mid Z) = 0$. Indeed, by the law of iterated expectations (LIE),

$$\begin{aligned}\mathrm{ICM}(U \mid Z) &= \mathbb{E}\left\{K(Z, Z^\dagger)(U^\dagger - \mathbb{E}U)\mathbb{E}[(U - \mathbb{E}U) \mid Z, Z^\dagger, U^\dagger]\right\}\\ &= \mathbb{E}\left\{K(Z, Z^\dagger)(\underbrace{\mathbb{E}[U \mid Z] - \mathbb{E}U}_{=0 \, a.s. \text{ under } \mathbb{H}_o})(U^\dagger - \mathbb{E}U)\right\}.\end{aligned} \tag{5}$$

Under $\mathbb{H}_o$, $\mathbb{E}[U \mid Z]$ is degenerate, that is, $\mathbb{E}[U \mid Z] - \mathbb{E}U = 0 \ a.s.$ This is the first-order degeneracy problem in U/V-statistic-based estimator for (5) that accounts for the non-pivotal limiting distribution in (3). In this regard, we propose to replace the role of U in the ICM metric with a random vector $V \in \mathbb{R}^{p_V}$ constructed such that $\mathbb{E}[V \mid Z]$ is non-degenerate under both the null and alternative hypotheses while preserving the omnibus property (2). In other words, we propose to measure the mean independence of U on Z with the assistance of a random vector V by

$$\delta(V) := \mathbb{E}\big[K(Z, Z^\dagger)(U^\dagger - \mathbb{E}U)(V - \mathbb{E}V)\big].$$

The resulting test relies on the empirical estimator of $\delta(V)$.

The two key properties—*first-order non-degeneracy* and *omnibusness*—ensure a pivotal limiting distribution under the null and the consistency of the test, respectively. Specifically, we require (1) V to be mean dependent on Z to rule out first-order degeneracy and (2) V to be chosen such that $\delta(V) \neq 0$ under $\mathbb{H}_a$. The first condition is easy to satisfy by selecting V to be a measurable function of Z. In contrast, the second condition is non-trivial: randomly selecting V may lead to the same limitation as the classical CM test if it happens to be orthogonal to the direction of departure under $\mathbb{H}_a$. This trade-off mirrors the strengths and weaknesses of existing methods: while the CM test is first-order non-degenerate, it may lack consistency under $\mathbb{H}_a$; conversely, the ICM test is consistent under $\mathbb{H}_a$, but suffers from first-order degeneracy, resulting in a non-pivotal limiting distribution under $\mathbb{H}_o$.

To this end, the proposed testing procedure integrates the respective strengths of two existing approaches in the literature: the pivotality of the CM test and the omnibus property of the ICM test. In particular, we focus on the quantity

$$V_h = \big[h(Z),\ U - h(Z)\big]^\top, \tag{6}$$

https://doi.org/10.1017/S0266466626100437 Published online by Cambridge University Press

such that

$$\delta_h := \delta(V_h) = \begin{bmatrix} \mathbb{E}[K(Z,Z^\dagger)(h(Z)-\mathbb{E}[h(Z)])(U^\dagger-\mathbb{E}U)] \\ \mathrm{ICM}(U \mid Z) - \mathbb{E}[K(Z,Z^\dagger)(h(Z)-\mathbb{E}[h(Z)])(U^\dagger-\mathbb{E}U)] \end{bmatrix}. \tag{7}$$

The two components in V_h, namely, $h(Z)$ and $U-h(Z)$, serve different purposes. First, one can view $h(Z)$ as the assistant function where prior information about the alternative can be incorporated, as done in the classical CM test. Second, the inclusion of U in the second element draws inspiration from ICM tests, which gives the omnibus property of our test, that is, it has non-trivial power in all possible directions of the alternative. This structure acts as a safeguard in scenarios where the chosen assistant $h(Z)$ is (nearly) orthogonal to the alternative, a situation in which traditional CM tests suffer from trivial power. Moreover, the $-h(Z)$ in the second element ensures that δ_h is first-order non-degenerate and that the proposed test statistic is pivotal (χ^2-distributed) under the null hypothesis. We remark that the choice of V can be readily extended beyond two dimensions, although in this article, we focus on the specific two-dimensional case for clarity and analytical tractability.

The following fundamental result establishes the omnibus and first-order non-degeneracy properties of δ_h.

LEMMA 2.1. *Let $h(Z)$ denote an arbitrary measurable and non-degenerate function of Z, in the sense that $\mathbb{P}\big(h(Z) = \mathbb{E}[h(Z)]\big) < 1$. Then: (a) δ_h defined in (7) satisfies the omnibus property, that is, $\|\delta_h\| = 0$ if and only if (2) holds and (b) δ_h exhibits first-order non-degeneracy.*

Proof. Let $\delta_h^{(1)} = \mathbb{E}[K(Z,Z^\dagger)(h(Z)-\mathbb{E}[h(Z)])(U^\dagger-\mathbb{E}U)]$ and $\delta_h^{(2)} = \mathrm{ICM}(U \mid Z) - \delta_h^{(1)}$. Then, from (7), we have that $\delta_h = [\delta_h^{(1)}, \delta_h^{(2)}]^\top$ and $\delta_h^{(1)} + \delta_h^{(2)} = \mathrm{ICM}(U \mid Z) > 0$ under $\mathbb{H}_a$. This implies that under $\mathbb{H}_a$, at least one element in δ_h is strictly positive (hence strictly different from zero); δ_h satisfies the omnibus property. Note that $h(Z)$ is non-degenerate, and that under $\mathbb{H}_o$, $\mathbb{E}\big[V_h - \mathbb{E}[V_h] \mid Z\big] = \big[(h(Z) - \mathbb{E}[h(Z)]),\ \mathbb{E}[U \mid Z] - \mathbb{E}U - (h(Z) - \mathbb{E}[h(Z)])\big]^\top = (h(Z) - \mathbb{E}[h(Z)])[1, -1]^\top \neq 0$ with probability one. It follows that δ_h avoids first-order degeneracy. □

Lemma 2.1 implies that, with our construction of V_h in (6), δ_h is nonzero under the alternative. Therefore, $\|\delta_h\|$ can be interpreted as a metric for conditional mean independence, that is, $\|\delta_h\| = 0$ if and only if $\mathbb{E}[U \mid Z] = \mathbb{E}[U]$ *a.s.* While the performance of δ_h depends on the choice of $h(\cdot)$, we note that similar concerns arise in ICM tests, particularly regarding the selection of kernels. A practitioner who is agnostic about the alternative can use $h(Z) = \exp\big(Z^\top \mathbb{1}_{p_z}/\sqrt{p_z}\big)$, where $\mathbb{1}_p$ denotes a $p \times 1$ vector of ones.[1] The Maclaurin's series expansion, namely,

[1] The specification is appropriate when p_z is small. For larger values of p_z, it is advisable to de-mean each element of Z to ensure the variance of $h(Z)$ is robust to p_z.

https://doi.org/10.1017/S0266466626100437 Published online by Cambridge University Press

$\exp(z) = \sum_{l=0}^{\infty} z^l/l!$ shows the potential of $h(Z) = \exp\left(Z^\top \mathbb{1}_{p_z}/\sqrt{p_z}\right)$ to capture different directions under the alternative.[2]

A key message of Lemma 2.1 is that symmetrizing any non-degenerate measurable function $h(\cdot)$ of Z about U guarantees the omnibus and first-order non-degeneracy properties of δ_h. The requirement on $h(Z)$ in Lemma 2.1 can hardly be termed a "condition," as it imposes only measurability and non-degeneracy. The proof of Lemma 2.1 provides an insight into the inclusion of U linearly in the construction of V. This brings in the term $\mathrm{ICM}(U \mid Z)$ in δ_h that is strictly positive under $\mathbb{H}_a$ and contributes power under the alternative. The proposed test thus draws its consistency from the ICM metric.

Since $h(Z)$ can be chosen arbitrarily, the practitioner does not bear the burden of "carefully" selecting functions such as polynomials that provide power by approximating $\mathbb{E}[U \mid Z]$ under $\mathbb{H}_a$, as required in nonparametric specification tests (e.g., Wooldridge, 1992; Yatchew, 1992; Zheng, 1996). In addition, the user-specified $h(Z)$ provides extra flexibility as the practitioner can use it to augment the power of the test in given directions, unlike existing ICM-based bootstrap tests.

Remark 2.1. We assume that p_z is fixed in this study. In high-dimensional settings where $p_z \to \infty$, Zhang et al. (2018, Remark 2.2) demonstrate that ICM criteria only capture linear dependence. Our χ^2-test derives its consistency property and part of its power from the ICM metric so it cannot be expected to perform better in high dimensions. Recent studies have explored the challenge of testing independence between high-dimensional random vectors using distance covariance-based statistics (Székely, Rizzo, and Bakirov, 2007), revealing that the performance of such tests depends critically on the relative growth rates of the dimensions of the random variables and the sample size n (Zhu et al., 2020; Chakraborty and Zhang, 2021; Gao et al., 2021; Zhang, Zhang, and Zhou, 2023). Consequently, extending the current testing procedure to high dimensions requires a separate study that is deferred to future research.

2.3. Extension to Testing the Nullity of $\mathbb{E}[U \mid Z]$

In some applications, such as specification testing, one may be interested in the almost sure nullity of $\mathbb{E}[U \mid Z]$ directly, that is,

$$\mathbb{H}_o^* : \mathbb{E}[U \mid Z] = 0 \ a.s.; \quad \mathbb{H}_a^* : \mathbb{P}(\mathbb{E}[U \mid Z] = 0) < 1. \tag{8}$$

This is an augmented version of (4) which further imposes $\mathbb{E}[U] = 0$, that is, a joint hypothesis of conditional mean independence and nullity of the unconditional mean. To this end, we follow Su and Zheng (2017) by augmenting δ_h with an additional quantity that accounts for $\mathbb{E}[U] = 0$ under $\mathbb{H}_o^*$. In particular, one may

[2]The covariance matrix of Z ought to be positive definite, that is, $\mathbb{P}(Z^\top a = c) < 1 \ \forall a \in \mathbb{R}^{p_z}$ and $\forall c \in \mathbb{R}$, to avoid possible degeneracy in the linear combination $Z^\top \mathbb{1}_{p_z}$.

https://doi.org/10.1017/S0266466626100437 Published online by Cambridge University Press

consider the metric

$$\delta_h^* = \delta_h + \mathbb{E}|K(Z,Z^\dagger)|\mathbb{E}U\mathbb{E}\big[h(Z),\ U-h(Z))\big]^\top.$$

The following result shows that δ_h^* has the omnibus property with respect to (8).

LEMMA 2.2. $\delta_h^* = 0$ *if and only if* $\mathbb{E}[U \mid Z] = 0$ *a.s.*

To conserve space, the discussion below focuses on δ_h. Theoretical properties of the χ^2-test based on δ_h^* are presented in Section S.3 of the Supplementary Material.

3. TEST STATISTIC AND THEORETICAL PROPERTIES

In this section, we construct the test statistic based on δ_h. In Section 3.1, we present a framework that unifies the tests of mean independence and model specification. We analyze the asymptotic behavior of $\widehat{\delta}_h$, the empirical estimator for δ_h, and the χ^2-distributed test statistic under the null hypothesis, as well as under local and fixed alternatives in Sections 3.2 and 3.3, respectively. Section 3.4 uses Bahadur slopes to compare the power of the χ^2-test and bootstrap-based ICM tests.

3.1. A Unified Framework

The mean independence testing problem and the specification testing problem are two closely related problems in econometric analyses. Here, we present a unified framework for both tasks. Let $X \in \mathbb{R}^{p_x}$ and $Z \in \mathbb{R}^{p_z}$ be two random vectors. We consider an econometric model defined by the following CM restriction:

$$\mathbb{E}[U(X;\theta_o)|Z] = 0 \quad a.s., \tag{9}$$

for a unique model parameter $\theta_o \in \Theta \subset \mathbb{R}^k$. Here, $U : \mathbb{R}^{p_x} \times \Theta \mapsto \mathbb{R}^{p_u}$ is the econometric model that is assumed to be known, and Θ is the parameter space. Given *iid* observations $\{X_i, Z_i\}_{i=1}^n$, and an empirical estimator $\widehat{\theta}_n$ for θ_o, we are interested in assessing the model specification in (9).

The above framework encompasses a wide range of applications, including treatment effect analyses (e.g., Sant'Anna and Song, 2019; Callaway and Karami, 2023), Euler and Bellman equations (Hansen and Singleton, 1982; Escanciano, 2018), the hybrid New Keynesian Phillips curve (Choi, Escanciano, and Guo, 2022), forecast rationality (Hansen and Hodrick, 1980), and tests of conditional equal and predictive ability (Giacomini and White, 2006) (see Li et al., 2026 for a comprehensive discussion).

In the mean independence testing problem outlined in Section 2.2, the expressions for ICM$(U \mid Z)$ and δ_h involve the nuisance parameter $\mathbb{E}U$. In this case, we can simply set $X = U$, $\theta_o = \mathbb{E}U$, and $U(X;\theta) = X - \theta$, then testing for (9) reduces to the hypothesis (4). In the class of nonlinear models with additively separable errors $X^{(1)} = g(X^{(-1)};\beta_o) + U$, where $X^{(1)}$ denotes the first element of X, $X^{(-1)}$ is the sub-vector of X that excludes $X^{(1)}$, and U is the model error (up to a location

https://doi.org/10.1017/S0266466626100437 Published online by Cambridge University Press

shift), we let

$$U(X;\theta) = X^{(1)} - g(X^{(-1)};\beta) - \theta_c,$$

where $\theta = [\theta_c,\ \beta^\top]^\top$, and θ_c is the location parameter such that $\mathbb{E}[X_1 - g(X_{-1};\beta) - U] = \theta_c$. Including the location parameter θ_c in θ becomes necessary as the ICM metric and δ_h can only identify mean independence up to an unknown nuisance mean shift. Nullity testing is discussed in Section 2.3.

In practice, a natural estimator for δ_h in (7) is given by

$$\widehat{\delta}_h = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} K(Z_i, Z_j) U(X_i; \widehat{\theta}_n)$$
$$\left[(h(Z_j) - \mathbb{E}_n[h(Z)]),\ U(X_j;\widehat{\theta}_n) - (h(Z_j) - \mathbb{E}_n[h(Z)]) \right]^\top. \tag{10}$$

We remark that, for the mean independence testing problem in (4), $U(X_i;\widehat{\theta}_n) = U_i - \widehat{\theta}_n$ with $\widehat{\theta}_n = \mathbb{E}_n U$.

3.2. Asymptotic Distribution

In this section, we analyze the asymptotic behavior of $\widehat{\delta}_h$ under the null hypothesis, local alternatives, and fixed alternatives. Let $D := [U, X^\top, Z^\top]^\top$, and define $\widetilde{U} := U(X;\theta_o)$ such that $\mathbb{E}\widetilde{U} = 0$. We consider the following sequences of local alternatives: Pitman local alternatives $\mathbb{H}_{an} : \mathbb{E}[\widetilde{U} \mid Z] = n^{-1/2} a(Z)$ and milder (non-Pitman) local alternatives $\mathbb{H}'_{an} : \mathbb{E}[\widetilde{U} \mid Z] = n^{-1/4} a(Z)$, where $a(Z)$ is a non-degenerate measurable function of Z satisfying $\mathbb{E}[a(Z)] = 0$. For completeness, we impose the following regularity conditions.

Assumption 3.1. $\{D_i\}_{i=1}^n$ are independently and identically distributed (*iid*).

Assumption 3.2. $\mathbb{E}\left[K(Z, Z^\dagger)^4\right] + \mathbb{E}\left[U^4\right] + \mathbb{E}\left[h(Z)^4\right] < \infty$.

Assumption 3.3. $U(X;\theta)$ is differentiable in θ and admits the following first-order expansion:

$$U(X;\theta) = \widetilde{U} + \frac{\partial U(X;\bar{\theta})}{\partial \theta'}(\theta - \theta_o) := \widetilde{U} - G(X;\bar{\theta})^\top (\theta - \theta_o),$$

where $G(x;\theta)$ is continuous in θ for all x in the support of X, $\mathbb{E}\left[\|G(X;\theta)\|^2\right] < \infty$ for all θ, and $\bar{\theta}$ lies on the line segment between θ and θ_o, that is, $\|\bar{\theta} - \theta_o\| \leq \|\theta - \theta_o\|$.

Assumption 3.4. $\widehat{\theta}_n$ is a $\sqrt{n}$-consistent estimator of θ_o with the asymptotically linear representation

$$\sqrt{n}(\widehat{\theta}_n - \theta_o) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_{\theta,i} + o_p(1),$$

where $\xi_{\theta,i} \in \mathbb{R}^k$ satisfies $\mathbb{E}[\xi_{\theta,i}] = \mathbf{0}$ and $\mathbb{E}\left[\|\xi_{\theta,i}\|^2\right] < \infty$.

https://doi.org/10.1017/S0266466626100437 Published online by Cambridge University Press

The assumption of *iid* observations in Assumption 3.1 simplifies the theoretical analyses.[3] Assumption 3.2 is standard (e.g., Serfling, 1980, Sect. 5.5.1 Thm. A); it is needed to establish the asymptotic normality of $\sqrt{n}(\widehat{\delta}_h - \delta_h)$. Although the moment restriction on U cannot be relaxed, that of $K(\cdot)$ can be relaxed through the choice of bounded kernels or suitable transformations of Z. Assumption 3.3 regulates the smoothness and moment conditions of the possibly nonlinear parameterized error $U(X;\theta)$. Assumption 3.4 is akin to Escanciano (2009a, Assump. A3); it assumes that the estimator $\widehat{\theta}_n$ is consistent with respect to its probability limit θ_o. The Bahadur linear representation holds for a broad class of estimators, including commonly used estimators, such as (nonlinear) least squares, maximum likelihood, and the (generalized) method of moments, under standard regularity conditions.

The result on the limiting behavior of the proposed test statistic relies on the following asymptotically linear representation of $\widehat{\delta}_h$ in the following lemma. Define $\widetilde{h}(Z) := h(Z) - \mathbb{E}h(Z)$.

LEMMA 3.1. *Under Assumptions 3.1–3.4, $\widehat{\delta}_h$ has the following asymptotically linear representation:*

$$\sqrt{n}\big(\widehat{\delta}_h - \delta_h\big) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_h(D_i) + o_p(1),$$

where $\xi_h(D_i) := 2\big[\xi_{h,1}(D_i),\ \xi_{h,2}(D_i) - \xi_{h,1}(D_i)\big]^\top$, $\delta_h = \big[\delta_h^{(1)}, \delta_h^{(2)}\big]^\top$,

$$\xi_{h,1}(D_i) := \psi_h^{(1)}(D_i) - \delta_h^{(1)} - H_1^\top[\xi_{\theta,i}^\top, \widetilde{h}(Z_i)]^\top,$$

$$\xi_{h,2}(D_i) := \psi_U^{(1)}(D_i) - \mathrm{ICM}(U \mid Z) - H_2^\top\xi_{\theta,i},$$

$$\psi_h^{(1)}(D_i) := \mathbb{E}[\psi_h(D_i, D_j)|D_i],$$

$$\psi_U^{(1)}(D_i) := \mathbb{E}[\psi_U(D_i, D_j)|D_i],$$

$\delta_h^{(1)} = \mathbb{E}[K(Z, Z^\dagger)\widetilde{h}(Z)\widetilde{U}^\dagger]$, $\delta_h^{(2)} = \mathrm{ICM}(U \mid Z) - \delta_h^{(1)}$, $\psi_h(D_i, D_j) := K(Z_i, Z_j)\big\{\widetilde{h}(Z_j)\widetilde{U}_i + \widetilde{h}(Z_i)\widetilde{U}_j\big\}/2$, $\psi_U(D_i, D_j) := K(Z_i, Z_j)\widetilde{U}_i\widetilde{U}_j$, $H_1 = (1/2)\Big[\mathbb{E}\big[K(Z, Z^\dagger)\widetilde{h}(Z^\dagger)G(X;\theta_o)\big]^\top, \mathbb{E}\big[K(Z, Z^\dagger)\widetilde{U}\big]\Big]^\top$, *and* $H_2 = \mathbb{E}\big[K(Z, Z^\dagger)\widetilde{U}^\dagger G(X;\theta_o)\big]$.

The proof of Lemma 3.1 exploits the classical Hoeffding decomposition (Serfling, 1980) for U-statistics and provides intuition for the two different rates of local alternatives. If the user-defined assistant $h(Z)$ is non-orthogonal to the alternative direction $\mathbb{E}\big[a(Z^\dagger)K(Z, Z^\dagger) \mid Z\big]$ under the Pitman local alternatives

[3] Our results are extensible to weak temporal dependence, panel data, and clustered data settings, but this lies beyond the scope of the current article.

https://doi.org/10.1017/S0266466626100437 Published online by Cambridge University Press

$\mathbb{H}_{an}: \mathbb{E}[\widetilde{U}|Z] = n^{-1/2}a(Z)$, then $\sqrt{n}\delta_h^{(1)}$ is non-diminishing, that is,

$$\sqrt{n}\delta_h^{(1)} = \mathbb{E}\big[K(Z,Z^\dagger)\widetilde{h}(Z)\sqrt{n}\mathbb{E}(\widetilde{U}^\dagger \mid Z^\dagger)\big] = \mathbb{E}[K(Z,Z^\dagger)\widetilde{h}(Z)a(Z^\dagger)] \not\to 0,$$

which drives the power in a manner similar to the classical CM test. If instead $h(Z)$ is (nearly) orthogonal to the alternative with $\sqrt{n}\delta_h^{(1)} \to 0$, then for the milder local alternatives $\mathbb{H}'_{an}: \mathbb{E}[\widetilde{U}|Z] = n^{-1/4}a(Z)$,

$$\begin{aligned}\sqrt{n}\delta_h^{(2)} \to \sqrt{n}\mathrm{ICM}(U \mid Z) &= \sqrt{n}\mathbb{E}[K(Z,Z^\dagger)\mathbb{E}(\widetilde{U}|Z)\mathbb{E}(\widetilde{U}^\dagger|Z^\dagger)] \\ &= \mathbb{E}[K(Z,Z^\dagger)a(Z)a(Z^\dagger)] > 0.\end{aligned}$$

In this case, the classical ICM test drives the power. As such, it provides robustness to the user's choice of direction, though this comes at the cost of a slower detection rate under local alternatives.

To study the limiting behavior of $\sqrt{n}(\widehat{\delta}_h - \delta_h)$, we define

$$\Omega_h := \mathrm{Var}[\xi_h(D_i)].$$

We distinguish $\Omega_{h,o}$ and $\Omega_{h,a}$, corresponding to specific expressions of Ω_h under $\mathbb{H}_o$ and $\mathbb{H}_a$, respectively. Both are generically referred to as Ω_h whenever the distinction is not necessary. In particular, under $\mathbb{H}_o$, $\psi_U^{(1)}(D_i) = \mathrm{ICM}(U \mid Z) = H_2^\top \xi_{\theta,i} = 0\ a.s.$ by the LIE, and $\sqrt{n}\big(\widehat{\delta}_h - \delta_h\big)$ reduces to $\sqrt{n}\widehat{\delta}_h = [1,-1]^\top \frac{2}{\sqrt{n}} \sum_{i=1}^n \xi_{h,1}(D_i) + o_p(1)$, with

$$\Omega_{h,o} := 4\mathbb{E}\big[\xi_{h,1}(D)^2\big] \times \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}. \tag{11}$$

The following result builds upon Lemma 3.1.

THEOREM 3.1. *Suppose Assumptions 3.1–3.4 hold, then*

(i) *under* $\mathbb{H}_o$, $\sqrt{n}\widehat{\delta}_h \xrightarrow{d} \mathcal{N}(0, \Omega_{h,o})$;
(ii) *under* $\mathbb{H}_{an}$, $\sqrt{n}\widehat{\delta}_h \xrightarrow{d} \mathcal{N}(a_o, \Omega_{h,o})$;
(iii) *under* $\mathbb{H}'_{an}$, $\sqrt{n}\widehat{\delta}_h \xrightarrow{d} \mathcal{N}(a'_o, \Omega_{h,o})$ *if* $\mathbb{E}[\widetilde{h}(Z)a(Z^\dagger)K(Z,Z^\dagger)] = 0$; *and*
(iv) *under* $\mathbb{H}_a$, $\sqrt{n}(\widehat{\delta}_h - \delta_h) \xrightarrow{d} \mathcal{N}(0, \Omega_{h,a})$;

where $a_o := [1,-1]^\top \mathbb{E}[\widetilde{h}(Z)a(Z^\dagger)K(Z,Z^\dagger)]$ *and* $a'_o := \big[0, \mathrm{ICM}\big(a(Z) \mid Z\big)\big]^\top$.

Theorem 3.1 offers a comprehensive asymptotic analysis of $\widehat{\delta}_h$ under the null, local alternatives, and fixed alternatives. In particular, part (i) shows that $\sqrt{n}\widehat{\delta}_h$ is non-degenerate and converges in distribution, which forms the basis of our χ^2-distributed test. Under the Pitman local alternatives $\mathbb{H}_{an}$, the test is only non-trivial in directions where the user-defined assistant $h(Z)$ is non-orthogonal to $\mathbb{E}\big[a(Z^\dagger)K(Z,Z^\dagger) \mid Z\big]$, that is, $\mathbb{E}[\widetilde{h}(Z)a(Z^\dagger)K(Z,Z^\dagger)] \neq 0$. This is expected, as the use of the assistant $h(Z)$ draws inspiration from classical CM tests. Part (iii) of Theorem 3.1 partially resolves this limitation: the local power becomes non-trivial in all directions due to the presence of $\mathrm{ICM}\big(a(Z) \mid Z\big)$ in the term $a'_o :=$

https://doi.org/10.1017/S0266466626100437 Published online by Cambridge University Press

$[0,\ \mathrm{ICM}\big(a(Z) \mid Z\big)]^\top$ under a sequence of alternatives converging to the null at the $n^{-1/4}$ rate. In other words, the test becomes omnibus for alternative directions that converge to the null at a rate slower than $n^{-1/4}$, thereby enhancing its ability to detect a broader class of local alternatives. The two types of local alternatives stem from our novel construction of the test, which combines components that operate at two different convergence rates. Part (iv) suggests that the resulting test is consistent under the fixed alternative.

Remark 3.1. The power performance of the proposed χ^2-test under the Pitman local alternative $\mathbb{H}_{an}$ is sensitive to the choice of $h(Z)$ and the ICM kernel $K(Z, Z^\dagger)$. A potential way to enhance the local power of the proposed test is to extend the estimand δ_h to a vector $\left\{\delta_{h,K}\right\}_{\{h,K\}\in\mathcal{H}\times\mathcal{K}}$, where the subscript in $\delta_{h,K}$ emphasizes the dependence of the parameter on both the function h and the ICM kernel K. Here, $\mathcal{H}$ and $\mathcal{K}$ are finite sets of measurable, non-degenerate functions of Z and valid ICM kernels, respectively.[4]

3.3. Test Statistic

Theorem 3.1 justifies the asymptotic normality of $\widehat{\delta}_h$, even under $\mathbb{H}_o$, which naturally motivates the Wald test statistic:

$$\widetilde{T}_{V,n} = n\widehat{\delta}_h^\top \widetilde{\Omega}_{h,n}^{-1}\widehat{\delta}_h, \tag{12}$$

where $\widetilde{\Omega}_{h,n}$ is a consistent estimator of Ω_h under both $\mathbb{H}_o$ and $\mathbb{H}_a$. This testing procedure is valid when Ω_h is positive definite. However, from (11), Ω_h is singular under $\mathbb{H}_o$ with $\mathrm{rank}(\Omega_{h,o}) = 1$.

Although it is tempting to replace the inverse matrix $\widetilde{\Omega}_{h,n}^{-1}$ in (12) with a generalized inverse matrix $\widetilde{\Omega}_{h,n}^{-}$, for example, the Moore–Penrose inverse, the resulting Wald statistic may still not have an asymptotic χ^2 distribution unless the rank condition, $\mathbb{P}\big(\mathrm{rank}(\widetilde{\Omega}_{h,n}) = \mathrm{rank}(\Omega_h)\big) \to 1 \quad \text{as } n \to \infty$ is satisfied (Andrews, 1987). To ensure that the rank condition is satisfied and thereby address this problem, we adopt the thresholding technique described in Lütkepohl and Burda (1997) (see also Duchesne and Francq, 2015; Dufour and Valéry, 2016). Let

$$\widetilde{\Omega}_{h,n} := \frac{1}{n-1}\sum_{i=1}^{n}\widehat{\xi}_h(D_i)\widehat{\xi}_h(D_i)^\top$$

be a consistent estimator of Ω_h (Sen, 1960), where $\widehat{\xi}_h(D_i)$ denotes the sample analog of $\xi_h(D_i)$ with parameters replaced by estimates.

By the singular value decomposition,

$$\widetilde{\Omega}_{h,n} = \widetilde{\Gamma}_n\widetilde{\Lambda}_n\widetilde{\Gamma}_n^\top, \tag{13}$$

[4]Another possibility is to extend V_h with a vector comprising a dictionary of transformations, for example, polynomials of Z (see Section S.5.3 of the Supplementary Material for some simulation evidence). A treatment of these extensions is omitted for considerations of space and scope.

https://doi.org/10.1017/S0266466626100437 Published online by Cambridge University Press

where $\widetilde{\Lambda}_n = \mathrm{diag}(\widetilde{\lambda}_1, \widetilde{\lambda}_2)$ is the diagonal matrix comprising the eigenvalues $\widetilde{\lambda}_1 \geq \widetilde{\lambda}_2 \geq 0$ of $\widetilde{\Omega}_{h,n}$, and the columns of $\widetilde{\Gamma}_n$ are the corresponding eigenvectors. Let $c_n = Cn^{-1/2+\iota}$, where $\iota \in (0, 1/2)$ is a small positive constant and $C \in (0, \infty)$ is a fixed constant, as in Dufour and Valéry (2016). Since $\mathrm{rank}(\Omega_h) \geq 1$ regardless of whether the null hypothesis $\mathbb{H}_o$ holds, we define the regularized estimator of Ω_h as

$$\widehat{\Omega}_{h,n} := \widetilde{\Gamma}_n \widehat{\Lambda}_{n,c_n} \widetilde{\Gamma}_n^\top, \quad \text{where} \quad \widehat{\Lambda}_{n,c_n} = \mathrm{diag}\big(\widetilde{\lambda}_1, \widetilde{\lambda}_2 \mathbf{1}(\widetilde{\lambda}_2 > c_n)\big). \tag{14}$$

The corresponding Moore–Penrose inverse is defined as

$$\widehat{\Omega}^-_{h,n} := \widetilde{\Gamma}_n \widehat{\Lambda}^-_{n,c_n} \widetilde{\Gamma}_n^\top, \quad \text{where} \quad \widehat{\Lambda}^-_{n,c_n} = \mathrm{diag}\big(\widetilde{\lambda}_1^{-1}, \widetilde{\lambda}_2^{-1} \mathbf{1}(\widetilde{\lambda}_2 > c_n)\big).$$

Finally, we define the regularized Wald test statistic:

$$T_{h,n} := n\widehat{\delta}_h^\top \widehat{\Omega}^-_{h,n} \widehat{\delta}_h. \tag{15}$$

In practice, $\mathbb{H}_o$ is rejected when $T_{h,n} > \chi^2_{1,1-\alpha}$, where $\chi^2_{1,1-\alpha}$ denotes the $1-\alpha$ quantile of the χ^2_1 distribution at a pre-specified significance level $\alpha \in (0,1)$. We remark that the ranks of $\Omega_{h,o}$ and $\Omega_{h,a}$ can differ: $\mathrm{rank}(\Omega_{h,o}) = 1$, while $\mathrm{rank}(\Omega_{h,a}) = 2$ if $\mathbb{P}(a(Z) = c \cdot \widetilde{h}(Z)) < 1$ for any $c \in \mathbb{R}$, and $\mathrm{rank}(\Omega_{h,a}) = 1$ otherwise. We use $c_n = \widetilde{\lambda}_1 n^{-1/3}$ following Lütkepohl and Burda (1997) and Dufour and Valéry (2016).[5]

Remark 3.2. $\mathrm{rank}(\Omega_{h,o}) = 1$ and $\mathrm{ICM}_n(U \mid Z) = O_p(n^{-1})$ under $\mathbb{H}_o$. From the proof of part (i) of Theorem 3.1, $\sqrt{n}\widehat{\delta}_h = \sqrt{n}\widehat{\delta}_h^{(1)}[1, \ -1]^\top + O_p(n^{-1/2})$, and

$$T_{h,n} = n\widehat{\delta}_h^\top \widehat{\Omega}^-_{h,n} \widehat{\delta}_h = \left(\frac{\widehat{\delta}_h^{(1)}}{\sqrt{\mathrm{Var}(\widehat{\delta}_h^{(1)})}} \right)^2 + o_p(1) \xrightarrow{d} \chi^2_1 \quad \text{as} \quad n \to \infty$$

under $\mathbb{H}_o$. This implies that $\sqrt{T_{h,n}}$ converges in distribution to the half standard normal under $\mathbb{H}_o$ and shares the interpretability (without a formal hypothesis test) of a two-sided t-test.

The next theorem justifies using (15) under the null, local alternative, and fixed alternative hypotheses.

THEOREM 3.2. *If* $\mathbb{E}|\xi_h(D)|^{4+\varepsilon} < \infty$ *for some* $\varepsilon > 0$, *and let* $c_n = Cn^{-1/2+\iota}$ *for some constants* $\iota \in (0, 1/2)$ *and* $C \in (0, \infty)$ *independent of n, then*

(i) *under* $\mathbb{H}_o$, $\widehat{\Omega}^-_{h,n} \xrightarrow{p} \Omega^-_{h,o}$, *and*

$$T_{h,n} \xrightarrow{d} \chi^2_1;$$

[5] For the robustness of the χ^2-test to variations of $c_n = \widetilde{\lambda}_1 n^{-\iota}$, $\iota \in (0, 1/2)$, and other common truncation criteria, see Section S.5.4 of the Supplementary Material.

https://doi.org/10.1017/S0266466626100437 Published online by Cambridge University Press

(ii) *under* $\mathbb{H}_{an}$, $\widehat{\Omega}^{-}_{h,n} \xrightarrow{p} \Omega^{-}_{h,o}$, *and the asymptotic local power is given by*

$$\lim_{n\to\infty} \mathbb{P}(T_{h,n} > \chi^2_{1,1-\alpha}) = \mathbb{P}\left(\chi^2_1(b_o) > \chi^2_{1,1-\alpha}\right),$$

where $b_o := a_o^\top \Omega^{-}_{h,o} a_o = \dfrac{(\mathbb{E}[\widetilde{h}(Z)a(Z^\dagger)K(Z,Z^\dagger)])^2}{4\mathbb{E}[\xi_{h,1}(D)^2]}$, a_o *is defined in Theorem 3.1, and* $\chi^2_1(b_o)$ *is a non-central* χ^2_1 *random variable;*

(iii) *under* $\mathbb{H}'_{an}$, $\widehat{\Omega}^{-}_{h,n} \xrightarrow{p} \Omega^{-}_{h,o}$, *and the asymptotic local power is given by*

$$\lim_{n\to\infty} \mathbb{P}(T_{h,n} > \chi^2_{1,1-\alpha}) = \mathbb{P}\left(\chi^2_1(b'_o) > \chi^2_{1,1-\alpha}\right),$$

if $\mathbb{E}[\widetilde{h}(Z)a(Z^\dagger)K(Z,Z^\dagger)] = 0$, *where* $b'_o := a_o^{'\top} \Omega^{-}_{h,o} a'_o = \dfrac{\mathrm{ICM}^2\big(a(Z) \mid Z\big)}{16\mathbb{E}[\xi_{h,1}(D)^2]} > 0$, *and* a'_o *is defined in Theorem 3.1; and*

(iv) *under* $\mathbb{H}_a$, $\widehat{\Omega}^{-}_{h,n} \xrightarrow{p} \Omega^{-}_{h,a}$, *and if* $\delta_h \notin \mathcal{M}_0$, *where* $\mathcal{M}_0$ *is the eigenspace associated with the null eigenvalue of* $\Omega_{h,a}$,

$$\lim_{n\to\infty} \mathbb{P}(T_{h,n} > \chi^2_{1,1-\alpha}) = 1.$$

Theorem 3.2 establishes the consistency of the thresholded estimator of the Moore–Penrose inverse of the covariance matrix and characterizes the asymptotic behavior of the resulting test statistic, in parallel with the analysis in Theorem 3.1. In particular, under the local alternatives $\mathbb{H}_{an}$ of part (ii), when $a_0 = \mathbb{E}[\widetilde{h}(Z)a(Z^\dagger)K(Z,Z^\dagger)] = 0$, the test exhibits trivial power because $b_o = 0$. However, this limitation is partially remedied in part (iii), where the modified limiting direction b'_o remains positive.

Remark 3.3. Under $\mathbb{H}_a$, $\delta_h \notin \mathcal{M}_0$ always holds for the mean independence test. The proof is provided in Lemma S.2.1 in the Supplementary Material.

3.4. Power Comparison with the Bootstrap-Based ICM Test

In this section, we examine the efficiency of the proposed test statistic compared with the bootstrap-based ICM test by adopting the approach of Bahadur (1960). Under a fixed alternative, both tests demonstrate consistency as n tends to infinity. The Bahadur slope in Bahadur (1960) allows us to further compare the rate of convergence of p-values to zero as n increases.

Let $S_G(t) := \mathbb{P}\Big(\sum_{k=1}^{\infty} \lambda_k G_k^2 > t\Big)$ and $S_T(t) := \mathbb{P}\big(\chi^2_1 > t\big)$ be the survival functions of the asymptotic null distributions of the ICM test statistic $n\mathrm{ICM}_n(U \mid Z)$, and the pivotal test statistic, $T_{h,n}$, respectively. Their Bahadur slopes are, respectively, given by

$$c_G = \lim_{n\to\infty} -\frac{2}{n} \log S_G(n\mathrm{ICM}_n(U \mid Z)) \quad \text{and} \quad c_T = \lim_{n\to\infty} -\frac{2}{n} \log S_G(T_{h,n}).$$

https://doi.org/10.1017/S0266466626100437 Published online by Cambridge University Press

THEOREM 3.3. *Suppose* $\widehat{\Omega}^{-}_{h,n} \xrightarrow{p} \Omega^{-}_{h,a}$, *and the conditions of Theorem 3.2 hold, then under* $\mathbb{H}_a$, *the (approximate) Bahadur slopes of the ICM test statistic* $n\text{ICM}_n(U \mid Z)$ *and the pivotal test statistic* $T_{h,n}$ *are, respectively, given by*

$$c_G = \frac{\text{ICM}(U \mid Z)}{\lambda_1} \quad \text{and } c_T = \delta_a^\top \Omega^{-}_{h,a} \delta_a,$$

where λ_1 *is the leading eigenvalue associated with the limiting null distribution in (3), and*

$$\delta_a = \left[\mathbb{E}\big[\widetilde{h}(Z)a(Z^\dagger)K(Z,Z^\dagger)\big],\ \text{ICM}(U \mid Z) - \mathbb{E}\big[\widetilde{h}(Z)a(Z^\dagger)K(Z,Z^\dagger)\big]\right]^\top.$$

The approximate Bahadur slopes presented in Theorem 3.3 are primarily of theoretical interest. Conducting a comprehensive comparison of these slopes is challenging as they depend on data-dependent quantities, such as λ_1, $a(Z)$, and user-specified variables, such as $h(Z)$ and the kernel $K(\cdot,\cdot)$.

Remark 3.4. The main implication of Theorem 3.3 is that, even with a fixed ICM kernel, neither test is uniformly more powerful across all data-generating processes (DGPs) satisfying Assumption 3.2. This limitation arises because the true alternative is unknown and the function $h(\cdot)$ is user-specified. Table S.1 in the Supplementary Material presents a numerical illustration of this result.

4. MONTE CARLO EXPERIMENTS—SPECIFICATION TEST

This section examines the size and power of the proposed χ^2 specification test compared to bootstrap-based ICM procedures via simulations. The specification of V is $\widehat{V}_h = [h(Z),\ \widehat{U} - h(Z)]^\top$ with $h(Z) = \exp\left(Z^\top \mathbb{1}_{p_z}/\sqrt{p_z}\right)$. For the proposed χ^2-test, the regularized inverse in (12) is computed using $c_n = \widetilde{\lambda}_1 n^{-1/3}$, where $\widetilde{\lambda}_1$ is the leading eigenvalue of $\widetilde{\Omega}_{h,n}$. We consider two commonly used ICM kernels: the Gaussian $K(Z,Z^\dagger) = \exp\left(-0.5\|Z - Z^\dagger\|^2\right)$ (e.g., Bierens, 1982) and the negative Euclidean $K(Z,Z^\dagger) = -\|Z - Z^\dagger\|$ (Shao and Zhang, 2014). For a given ICM kernel, the proposed χ^2-test is compared to ICM tests based on the MB of Escanciano (2024) and the standard wild bootstrap (e.g., Escanciano, 2006a; Su and Zheng, 2017). The empirical size and power curves are based on 1,000 Monte Carlo replicates. Nine hundred ninety-nine bootstrap samples are used for the bootstrap procedures.[6] Other simulation results on specification and mean independence tests are available in Section S.5 of the Supplementary Material.

[6]Both bootstrap procedures are conducted using the Mammen (1993) two-point distributed auxiliary variables.

https://doi.org/10.1017/S0266466626100437 Published online by Cambridge University Press

4.1. Specifications

We consider the linear model

$$Y = \theta_c + \sum_{l=1}^{5} X_l \theta_l + U,$$

with and without excluded instruments.[7] The DGPs are varied through U:

$$\text{LS1}: U = \frac{\mathcal{E}}{\sqrt{1+Z_1^2}};$$

$$\text{LS2}: U = \frac{\gamma}{5\sqrt{2}} \sum_{l=1}^{5} Z_l^2 + \frac{\mathcal{E}}{\sqrt{1+Z_1^2}};$$

$$\text{LS3}: U = \gamma \sum_{l=1}^{5} \frac{\cos(2Z_l)}{\sqrt{2(1-\exp(-8))}} + \frac{\mathcal{E}}{\sqrt{1+Z_1^2}}; \text{ and}$$

$$\text{LS4}: U = \gamma \sum_{l=1}^{5} \left(\exp(-Z_l^2/3) - \sqrt{3/5}\right) + \frac{\mathcal{E}}{\sqrt{1+Z_1^2}};$$

where $\mathcal{E} \sim \mathcal{N}(0,1)$ independently of Z which follows the multivariate normal distribution with mean zero and covariance matrix $\Sigma_{ll'} = 0.25^{|l-l'|}$, $[\theta_c, \theta_1, \ldots, \theta_5]^\top = \mathbb{1}_6$, and $\gamma \in [0,1]$ tunes the deviation away from $\mathbb{H}_0$. $X = Z$ in DGP LS1 and $X_1 = (1.5Z_1 + \widetilde{\mathcal{E}})/\sqrt{3.25}$ with $\widetilde{\mathcal{E}} \sim \mathcal{N}(0,1)$ and $\text{Cov}(\widetilde{\mathcal{E}}, \mathcal{E}) = 0.25$ in DGPs LS2 through LS4.[8] Thus, X is exogenous under DGP LS1 and endogenous under DGPs LS2 through LS4. LS1 is estimated via ordinary least squares (OLS) while the remaining DGPs are estimated using the IVs estimator with Z as the instrument. Heteroskedasticity of arbitrary form is imposed in all DGPs.

Different sample sizes $n \in \{200, 400, 600, 800\}$ serve to examine the empirical size and local power of the test using DGPs LS1 and LS2 (with $\gamma = 0$). For the analyses of power in DGPs LS2 through LS4, the sample size is kept at $n = 400$, while γ is varied in order to study the power of the proposed χ^2-test in comparison to the bootstrap-based ICM procedures.

4.2. Empirical Size and Power

Table 2 presents the empirical sizes corresponding to DGPs LS1 (under strictly exogenous X) and LS2 (under endogenous X instrumented with Z) in addition to the local power under LS2 at three nominal levels: 10%, 5%, and 1%. One observes comparably good size control of the proposed χ^2-test and the bootstrap-based

[7] Section S.6 of the Supplementary Material provides simulation results on the specification of propensity scores using the logit model.

[8] The nonparametric quantity $\mathbb{E}[X_1 \mid Z_1]$ is needed for the MB under endogeneity (see Escanciano, 2024, Sect. 5). It is estimated using a third-order polynomial.

https://doi.org/10.1017/S0266466626100437 Published online by Cambridge University Press

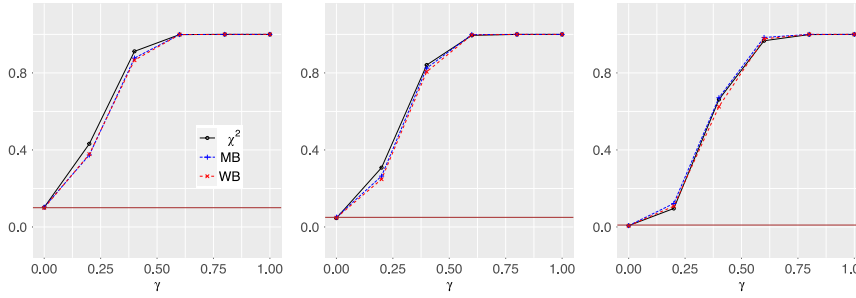
Table 2. Empirical size and local power

		10%			5%			1%		
n	Kernel	χ^2	MB	WB	χ^2	MB	WB	χ^2	MB	WB
LS1					Empirical size					
200	Gauss	0.096	0.103	0.103	0.053	0.049	0.052	0.007	0.012	0.012
	Euclid	0.101	0.072	0.057	0.048	0.031	0.021	0.008	0.003	0.003
400	Gauss	0.103	0.098	0.096	0.046	0.048	0.046	0.007	0.009	0.006
	Euclid	0.104	0.074	0.056	0.053	0.032	0.029	0.005	0.003	0.004
600	Gauss	0.089	0.102	0.106	0.053	0.050	0.047	0.009	0.011	0.008
	Euclid	0.100	0.084	0.085	0.047	0.036	0.034	0.006	0.006	0.005
800	Gauss	0.090	0.104	0.097	0.039	0.053	0.055	0.008	0.007	0.006
	Euclid	0.098	0.088	0.078	0.042	0.040	0.041	0.006	0.004	0.006
LS2					Empirical size					
200	Gauss	0.095	0.093	0.100	0.055	0.052	0.053	0.007	0.011	0.014
	Euclid	0.097	0.070	0.055	0.045	0.032	0.021	0.008	0.003	0.002
400	Gauss	0.102	0.103	0.099	0.046	0.051	0.048	0.007	0.007	0.006
	Euclid	0.103	0.079	0.056	0.056	0.031	0.028	0.005	0.005	0.003
600	Gauss	0.089	0.094	0.107	0.052	0.044	0.048	0.008	0.011	0.008
	Euclid	0.100	0.090	0.083	0.046	0.040	0.032	0.006	0.010	0.006
800	Gauss	0.090	0.099	0.095	0.039	0.051	0.055	0.009	0.008	0.006
	Euclid	0.099	0.089	0.079	0.042	0.041	0.037	0.006	0.004	0.006
LS2					Local power: $\gamma = 5/\sqrt{n}$					
200	Gauss	0.549	0.486	0.480	0.418	0.391	0.371	0.164	0.223	0.201
	Euclid	0.883	0.656	0.625	0.803	0.543	0.507	0.535	0.295	0.262
400	Gauss	0.605	0.545	0.522	0.459	0.413	0.390	0.202	0.226	0.182
	Euclid	0.887	0.707	0.689	0.813	0.601	0.575	0.588	0.376	0.341
600	Gauss	0.592	0.521	0.509	0.461	0.423	0.408	0.221	0.237	0.215
	Euclid	0.896	0.723	0.714	0.830	0.614	0.621	0.606	0.410	0.398
800	Gauss	0.606	0.547	0.539	0.473	0.441	0.428	0.243	0.246	0.225
	Euclid	0.912	0.743	0.742	0.840	0.651	0.644	0.632	0.420	0.422

procedures across all the sample sizes considered. This provides evidence in support of the proposed test's validity under both exogenous X and endogenous X with valid instruments Z.

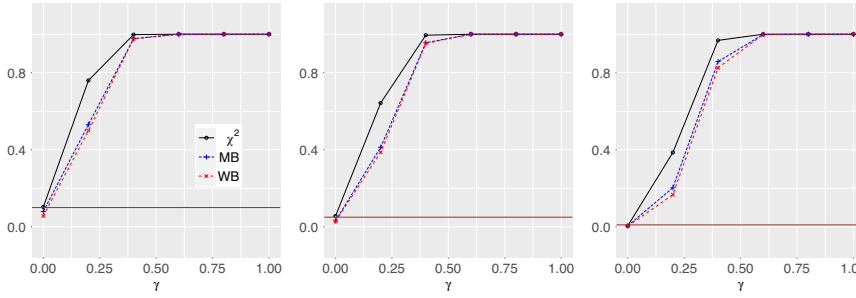
To ensure that the good size control of the χ^2-test in Table 2 is not achieved at the expense of power under the alternative, we consider its performance under both local and fixed alternatives. The results, presented in the third panel of Table 2, indicate that the local power of the proposed χ^2-test, along with the bootstrap-based procedures, is non-trivial. Figures 1–6 present power curves cor-

https://doi.org/10.1017/S0266466626100437 Published online by Cambridge University Press



(a) 10% (b) 5% (c) 1%

FIGURE 1. DGP LS2—Gaussian Kernel—$n = 400$.



(a) 10% (b) 5% (c) 1%

FIGURE 2. DGP LS2—Negative Euclidean—$n = 400$.

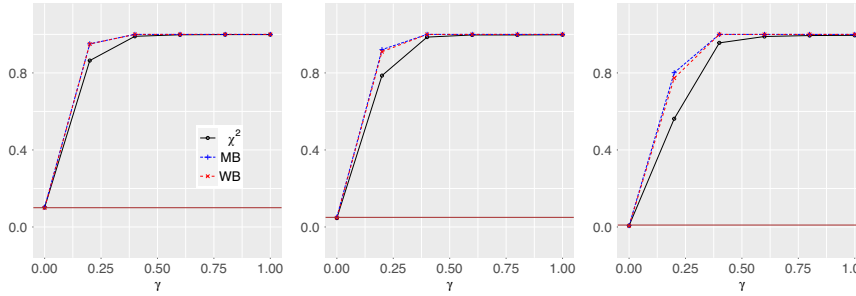
responding to DGPs LS2 through LS4 using the Gaussian and Negative Euclidean kernels. One observes that all tests demonstrate non-trivial power against the fixed alternative for variations of γ away from zero. The proposed test exhibits highly competitive power performance across all three DGPs and both kernels considered. Notably, the relative power of the χ^2-test compared to the bootstrap-based procedures demonstrates variability depending on the specific DGP and kernel, as highlighted in Theorem 3.3.

The χ^2-test proves useful when researchers seek a measure of correct model specification or mean independence without resorting to a formal hypothesis test, owing to the pivotality of its test statistic. In summary, our simulations demonstrate that the proposed χ^2-test, in addition to its enhanced interpretability compared to bootstrap-based ICM specification tests, exhibits good size control and comparable power performance.

4.3. Running Time

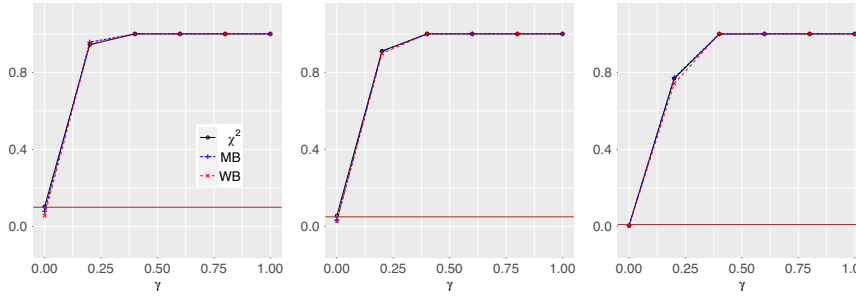
One key advantage of a pivotal test is its computational efficiency relative to bootstrap-based procedures, which becomes particularly critical in large samples. Accordingly, a comparison of the running times of competing tests is in

https://doi.org/10.1017/S0266466626100437 Published online by Cambridge University Press



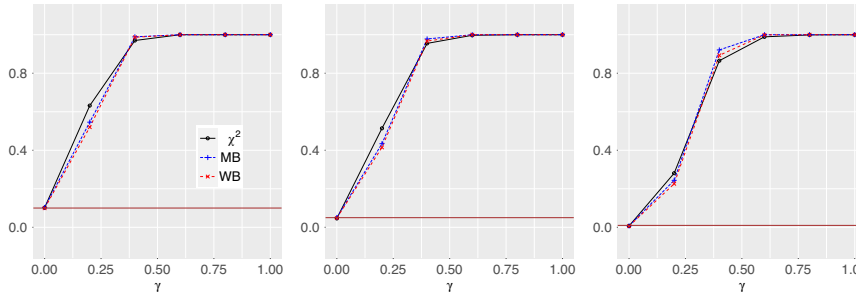
(a) 10% (b) 5% (c) 1%

FIGURE 3. DGP LS3—Gaussian Kernel—$n = 400$.



(a) 10% (b) 5% (c) 1%

FIGURE 4. DGP LS3—Negative Euclidean—$n = 400$.



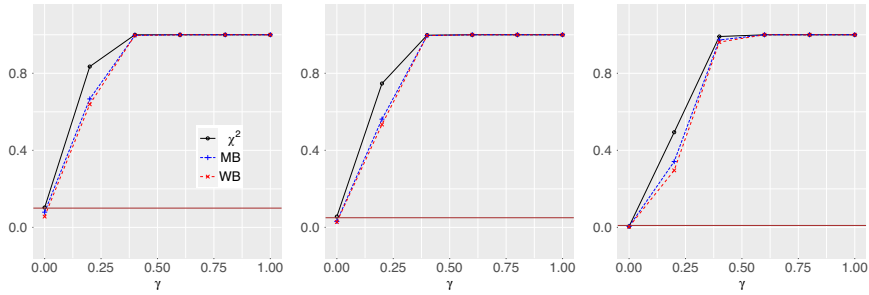
(a) 10% (b) 5% (c) 1%

FIGURE 5. DGP LS4—Gaussian Kernel—$n = 400$.

order. Table 3 presents a comparison between the proposed pivotal χ^2-test and the bootstrap-based ICM tests in terms of computational time for fixed ICM kernels.[9]

[9]Computations were performed on a 2.8 GHz Quad-Core Intel Core i7 with 16 GB of RAM, running on a MacBook Pro.

https://doi.org/10.1017/S0266466626100437 Published online by Cambridge University Press



(a) 10% (b) 5% (c) 1%

FIGURE 6. DGP LS4—Negative Euclidean—$n = 400$.

TABLE 3. Running time—specification test—DGP LS1

		Average running time (seconds)			Median relative time	
n	Kernel	χ^2	MB	WB	MB	WB
	Gauss	0.004	0.114	0.965	29.000	248.000
200		(0.002)	(0.021)	(0.063)	(16.833)	(142.208)
	Euclid	0.004	0.111	0.993	35.000	315.667
		(0.001)	(0.006)	(0.043)	(15.425)	(127.733)
	Gauss	0.011	0.440	1.100	47.222	116.422
400		(0.003)	(0.022)	(0.054)	(18.856)	(41.898)
	Euclid	0.009	0.435	1.095	58.786	145.571
		(0.002)	(0.016)	(0.055)	(23.858)	(55.071)
	Gauss	0.021	1.137	1.254	59.833	65.526
600		(0.004)	(0.035)	(0.048)	(18.407)	(18.479)
	Euclid	0.016	1.125	1.247	76.857	85.000
		(0.003)	(0.040)	(0.051)	(25.869)	(27.041)
	Gauss	0.038	2.427	1.552	65.859	41.789
800		(0.005)	(0.057)	(0.048)	(13.833)	(7.420)
	Euclid	0.030	2.471	1.584	87.554	55.930
		(0.004)	(0.053)	(0.040)	(22.298)	(13.505)

Note: The second row (in parentheses) for each sample size and kernel includes standard deviations and inter-quartile ranges for the average running times and relative times (the χ^2-test as the benchmark), respectively.

Table 3 illustrates the substantial computational advantage of employing the χ^2 specification test. The average running times for the χ^2-test are negligible compared to those of the bootstrap-based procedures across all considered sample sizes. A striking difference emerges in terms of median relative computational time

https://doi.org/10.1017/S0266466626100437 Published online by Cambridge University Press

(with the χ^2-test serving as a benchmark). While the MB procedure is generally faster than the wild bootstrap procedure, our findings indicate that the median relative computational time of the MB tends to increase with the sample size, whereas that of the wild bootstrap appears to decrease. At particularly large sample sizes ($n = 600$ and $n = 800$), the computational gain offered by the χ^2-test becomes considerable.

5. CONCLUSION

Despite the four-decade history of ICM tests, dating back to Bierens (1982) and encompassing numerous interesting contributions, a *bona fide* pivotalized ICM specification test remains lacking. This article achieves the objective of proposing an omnibus χ^2-test of specification and mean independence based on ICM metrics.

The proposed χ^2-test complements existing ICM tests by overcoming certain limitations of the latter. The test statistic can be constructed with functional forms that boost power in the direction of alternatives the researcher may have in mind. The test is computationally more efficient than commonly used bootstrap-based ICM tests and remains computationally viable even in large samples. In addition to providing a reliable pivotal test that derives its omnibus property from ICM metrics, the test statistic offers an easily interpretable metric of model specification and mean dependence, thereby obviating formal hypothesis tests.

In conclusion, we highlight several potential extensions to the current work. While this article offers a viable solution for pivotalizing ICM tests, it is noteworthy that this approach necessitates regularization under the null hypothesis to address the ill-posed inverse problem. Future research could explore alternative formulations of V that may yield solutions with more desirable theoretical properties. Furthermore, extending our analysis from the current fixed-dimensional setting to a high-dimensional framework presents an interesting avenue for research. Moreover, leveraging the projection approach proposed by Escanciano (2009b, 2024), which eliminates the estimation effect, could be advantageous in dealing with estimators whose limiting distributions are difficult to characterize, for example, the LASSO. Finally, extending our methodology to encompass time series data, clustered data, and multiple-equation models with (non)-smooth objective functions warrants future exploration.

COMPETING INTEREST STATEMENT

The authors declare that no competing interests exist.

FUNDING STATEMENT

F.J.'s research is supported in part by the National Key R&D Program of China (Grant No. 2024YFA1015700) and by the NSFC (Grant Nos. 72522009, 12201124, 12331009, 72271060, and 72432002).

https://doi.org/10.1017/S0266466626100437 Published online by Cambridge University Press

SUPPLEMENTARY MATERIAL

Proofs of all results and additional numerical results are provided in Jiang and Tsyawo (2025): Supplement to "A Consistent ICM-Based χ^2 Specification Test," Econometric Theory Supplementary Material. To view, please visit: https://doi.org/10.1017/S0266466626100437.

REFERENCES

Andrews, D. W. (1987). Asymptotic results for generalized Wald tests. *Econometric Theory*, 3(3), 348–358.

Antoine, B., & Lavergne, P. (2023). Identification-robust nonparametric inference in a linear IV model. *Journal of Econometrics*, 235(1), 1–24.

Bahadur, R. R. (1960). Stochastic comparison of tests. *The Annals of Mathematical Statistics*, 31(2), 276–295.

Bierens, H. J. (1982). Consistent model specification tests. *Journal of Econometrics*, 20(1), 105–134.

Bierens, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica*, 58, 1443–1458.

Bierens, H. J., & Ploberger, W. (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica*, 65, 1129–1151.

Callaway, B., & Karami, S. (2023). Treatment effects in interactive fixed effects models with a small number of time periods. *Journal of Econometrics*, 233(1), 184–208.

Chakraborty, S., & Zhang, X. (2021). A new framework for distance and kernel-based metrics in high dimensions. *Electronic Journal of Statistics*, 15(2), 5455–5522.

Choi, J., Escanciano, J. C., & Guo, J. (2022). Generalized band spectrum estimation with an application to the new Keynesian Phillips curve. *Journal of Applied Econometrics*, 37(5), 1055–1078.

De Jong, R. M. (1996). The Bierens test under data dependence. *Journal of Econometrics*, 72(1–2), 1–32.

Delgado, M. A. (1993). Testing the equality of nonparametric regression curves. *Statistics & Probability Letters*, 17(3), 199–204.

Delgado, M. A., Domínguez, M. A., & Lavergne, P. (2006). Consistent tests of conditional moment restrictions. *Annales d'Économie et de Statistique*, 81, 33–67.

Domínguez, M. A., & Lobato, I. N. (2004). Consistent estimation of models defined by conditional moment restrictions. *Econometrica*, 72(5), 1601–1615.

Domínguez, M. A., & Lobato, I. N. (2015). A simple omnibus overidentification specification test for time series econometric models. *Econometric Theory*, 31(4), 891–910.

Duchesne, P., & Francq, C. (2015). Multivariate hypothesis testing using generalized and {2} -inverses–with applications. *Statistics*, 49(3), 475–496.

Dufour, J.-M., & Valéry, P. (2016). Rank-robust Wald-type tests: A regularization approach. (Tech. rep.). Working Paper, McGill University/HEC.

Escanciano, J. C. (2006a). A consistent diagnostic test for regression models using projections. *Econometric Theory*, 22(6), 1030–1051.

Escanciano, J. C. (2006b). Goodness-of-fit tests for linear and nonlinear time series models. *Journal of the American Statistical Association*, 101(474), 531–541.

Escanciano, J. C. (2009a). On the lack of power of omnibus specification tests. *Econometric Theory*, 25(1), 162–194.

Escanciano, J. C. (2009b). Simple bootstrap tests for conditional moment restrictions. (Tech. rep.). Citeseer.

Escanciano, J. C. (2018). A simple and robust estimator for linear regression models with strictly exogenous instruments. *The Econometrics Journal*, 21(1), 36–54.

Escanciano, J. C. (2024). A gaussian process approach to model checks. *Annals of Statistics*, 52(5), 2456–2481.

https://doi.org/10.1017/S0266466626100437 Published online by Cambridge University Press

Fan, Y., & Li, Q. (2000). Consistent model specification tests: Kernel-based tests versus Bierens' ICM tests. *Econometric Theory*, 16(6), 1016–1041.

Gao, L., Fan, Y., Lv, J., & Shao, Q.-M. (2021). Asymptotic distributions of high-dimensional distance correlation inference. *Annals of Statistics*, 49(4), 1999.

Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6), 1545–1578.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, 64, 413–430.

Hansen, L. P., & Hodrick, R. J. (1980). Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. *Journal of Political Economy*, 88(5), 829–853.

Hansen, L. P., & Singleton, K. J. (1982). Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica*, 50, 1269–1286.

Hardle, W., & Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistics*, 21, 1926–1947.

Hong, Y., & White, H. (1995). Consistent specification testing via nonparametric series regression. *Econometrica*, 63, 1133–1159.

Kim, I., Balakrishnan, S., & Wasserman, L. (2020). Robust multivariate nonparametric tests via projection averaging. *Annals of Statistics*, 48(6), 3417–3441.

Li, J., Liao, Z., & Zhou, W. (2026). Learning before testing: A selective nonparametric test for conditional moment restrictions. *Review of Economics and Statistics*. 1–46.

Li, L., Ke, C., Yin, X., & Yu, Z. (2023). Generalized martingale difference divergence: Detecting conditional mean independence with applications in variable screening. *Computational Statistics & Data Analysis*, 180, 107618.

Li, Q., & Wang, S. (1998). A simple consistent bootstrap test for a parametric regression function. *Journal of Econometrics*, 87(1), 145–165.

Li, Y., & Song, X. (2022). Consistent test for conditional moment restriction models in reproducing kernel Hilbert spaces. Working Paper.

Li, Y., & Song, X. (2025). *A powerful chi-square specification test with support vectors*. Preprint, arXiv:2505.04414.

Lütkepohl, H., & Burda, M. M. (1997). Modified Wald tests under nonregular conditions. *Journal of Econometrics*, 78(2), 315–332.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Annals of Statistics*, 21, 255–285.

Newey, W. K. (1985). Maximum likelihood specification testing and conditional moment tests. *Econometrica*, 53, 1047–1070.

Raiola, A. (2024). Testing conditional moment restrictions: A partitioning approach. Working Paper. https://pure.au.dk/ws/portalfiles/portal/422827832/CMR.pdf

Sant'Anna, P. H., & Song, X. (2019). Specification tests for the propensity score. *Journal of Econometrics*, 210(2), 379–404.

Sen, P. K. (1960). On some convergence properties of U-statistics. *Calcutta Statistical Association Bulletin*, 10(1–2), 1–18.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. John Wiley & Sons.

Shao, X., & Zhang, J. (2014). Martingale difference correlation and its use in high-dimensional variable screening. *Journal of the American Statistical Association*, 109(507), 1302–1318.

Stute, W. (1997). Nonparametric model checks for regression. *Annals of Statistics*, 25, 613–641.

Su, L., & White, H. (2007). A consistent characteristic function-based test for conditional independence. *Journal of Econometrics*, 141(2), 807–834.

Su, L., & Zheng, X. (2017). A martingale-difference-divergence-based test for specification. *Economics Letters*, 156, 162–167.

Székely, G. J., Rizzo, M. L., & Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *Annals of Statistics*, 35(6), 2769–2794.

https://doi.org/10.1017/S0266466626100437 Published online by Cambridge University Press

Tauchen, G. (1985). Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics*, 30(1–2), 415–443.

Tsyawo, E. S. (2023). Feasible IV regression without excluded instruments. *The Econometrics Journal*, 26(2), 235–256.

Wooldridge, J. M. (1992). A test for functional form against nonparametric alternatives. *Econometric Theory*, 8(4), 452–475.

Yatchew, A. J. (1992). Nonparametric regression tests based on least squares. *Econometric Theory*, 8(4), 435–451.

Zhang, T., Zhang, Y., & Zhou, T. (2023). Statistical insights into HSIC in high dimensions. *Advances in Neural Information Processing Systems*, 36, 19145–19156.

Zhang, X., Yao, S., & Shao, X. (2018). Conditional mean and quantile dependence testing in high dimension. *Annals of Statistics*, 46(1), 219–246.

Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics*, 75(2), 263–289.

Zhu, C., Zhang, X., Yao, S., & Shao, X. (2020). Distance-based and RKHS-based dependence metrics in high dimension. *The Annals of Statistics*, 48(6), 3366–3394.

Zhu, L.-P., Li, L., Li, R., & Zhu, L.-X. (2011). Model-free feature screening for ultrahigh-dimensional data. *Journal of the American Statistical Association*, 106(496), 1464–1475.

https://doi.org/10.1017/S0266466626100437 Published online by Cambridge University Press