Supplement to "A Consistent ICM-based χ^2 Specification Test"

Feiyu Jiang*[1] and Emmanuel Selorm Tsyawo†[2]

[1]Department of Statistics and Data Science, Fudan University, Shanghai, China
[2]Department of Economics, Finance and Legal Studies, Culverhouse College of Business, University of Alabama

The supplementary material contains all technical proofs and additional simulation results. Section S.1 provides lemmata supporting proofs in Section S.2. Section S.3 studies the asymptotic properties of the χ^2-test based on δ_h^*. Section S.4 provides numerical comparisons of the Bahadur slopes discussed in Section 3.4. Section S.5 contains simulation results on mean independence tests, Section S.6 extends simulation results to specification testing of the non-linear logit model, and Section S.7 concludes by discussing the relation between the proposed χ^2-test and the family of CM tests.

S.1 Supporting Lemmata

By the analog and plug-in principles, the estimator of $\text{ICM}(U \mid Z)$ is given by

$$\text{ICM}_n(U \mid Z) := \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} K(Z_i, Z_j) U(X_i; \widehat{\theta}_n) U(X_j; \widehat{\theta}_n).$$

Lemma S.1.1. *Under Assumptions 3.1 to 3.4,* $\text{ICM}_n(U \mid Z)$ *has the following asymptotically linear representation:*

$$\sqrt{n}\big(\text{ICM}_n(U \mid Z) - \text{ICM}(U \mid Z)\big) = \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \Big\{ \psi_U^{(1)}(D_i) - \text{ICM}(U \mid Z) - H_2^\top \xi_{\theta,i} \Big\} + o_p(1).$$

Proof

---

*Email: jiangfy@fudan.edu.cn
†Corresponding Author. Email: estsyawo@gmail.com

Under Assumption 3.3 and the symmetry of the kernel $K(\cdot,\cdot)$,

$$\begin{aligned}
\mathrm{ICM}_n(U \mid Z) :=&\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} K(Z_i,Z_j)U(X_i;\widehat{\theta}_n)U(X_j;\widehat{\theta}_n)\\
=&\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i} K(Z_i,Z_j)\big(\widetilde{U}_i - G(X_i;\bar{\theta}_n)^\top(\widehat{\theta}_n-\theta_o)\big)\big(\widetilde{U}_j - G(X_j;\bar{\theta}_n)^\top(\widehat{\theta}_n-\theta_o)\big)\\
=&\frac{1}{n(n-1)}\sum_{i\neq j}^{n} K(Z_i,Z_j)\widetilde{U}_i\widetilde{U}_j\\
&+(\widehat{\theta}_n-\theta_o)^\top\left\{\frac{1}{n(n-1)}\sum_{i\neq j}^{n} K(Z_i,Z_j)G(X_i;\bar{\theta}_n)G(X_j;\bar{\theta}_n)^\top\right\}(\widehat{\theta}_n-\theta_o)\\
&-\frac{2}{n(n-1)}\sum_{i\neq j}^{n} K(Z_i,Z_j)\widetilde{U}_jG(X_i;\bar{\theta}_n)^\top(\widehat{\theta}_n-\theta_o)\\
:=&\sum_{k=1}^{3} G_k.
\end{aligned}$$

The first summand is a second-order U-statistic. Recall $\psi_U(D_i,D_j) := K(Z_i,Z_j)\widetilde{U}_i\widetilde{U}_j$, and $\psi_U^{(1)}(D_i) = \mathbb{E}\big[\psi_U(D_i,D_j)|D_i\big]$. By Assumption 3.1, Assumption 3.2, and the Hoeffding decomposition (e.g., Theorem 3 in section 1.3 of Lee (1990)),

$$\begin{aligned}
G_1 =&\frac{1}{n(n-1)}\sum_{i\neq j}^{n} K(Z_i,Z_j)\widetilde{U}_i\widetilde{U}_j\\
=&\binom{n}{2}^{-1}\sum_{i<j}^{n} K(Z_i,Z_j)\widetilde{U}_i\widetilde{U}_j\\
=&\mathrm{ICM}(U \mid Z) + \frac{2}{n}\sum_{i=1}^{n}\big[\psi_U^{(1)}(D_i) - \mathrm{ICM}(U \mid Z)\big] \qquad \text{(S.11)}\\
&+\binom{n}{2}^{-1}\sum_{i<j}^{n}\big[\psi_U(D_i,D_j) - \psi_U^{(1)}(D_i) - \psi_U^{(1)}(D_j) + \mathrm{ICM}(U \mid Z)\big].\\
=&\mathrm{ICM}(U \mid Z) + \frac{2}{n}\sum_{i=1}^{n}\big[\psi_U^{(1)}(D_i) - \mathrm{ICM}(U \mid Z)\big] + o_p(n^{-1/2}).
\end{aligned}$$

By Assumptions 3.1 and 3.4, $G_2 = O_p(n^{-1})$. For the third summand, note that by the law of large numbers for U-statistics, the continuous mapping theorem and Assumptions 3.1 to 3.3,

$$\frac{1}{n(n-1)}\sum_{i\neq j}^{n} K(Z_i,Z_j)\widetilde{U}_j G(X_i;\bar{\theta}_n)^\top \xrightarrow{p} \mathbb{E}[K(Z,Z^\dagger)\widetilde{U}^\dagger G(X;\theta_o)]^\top := H_2^\top.$$

Hence, $G_3 = -2H_2^\top \frac{1}{n}\sum_{i=1}^{n}\xi_{\theta,i} + o_p(n^{-1/2})$.

Combining all the terms above proves the assertion as claimed.

■

S.2 Proofs of results in the main text

S.2.1 Proof of Lemma 2.2

$$\begin{aligned}
\delta_h^* &= \Big[\mathbb{E}[\widetilde{U}^\dagger\widetilde{h}(Z)K(Z,Z^\dagger)],\ \mathrm{ICM}(U\mid Z)-\mathbb{E}[\widetilde{U}^\dagger\widetilde{h}(Z)K(Z,Z^\dagger)]\Big]^\top \\
&\quad + \big[\mathbb{E}[h(Z)]\mathbb{E}U,\ (\mathbb{E}U)^2-\mathbb{E}[h(Z)]\mathbb{E}U\big]^\top\,\mathbb{E}|K(Z,Z^\dagger)| \\
&= \Big[\mathbb{E}[\widetilde{U}^\dagger\widetilde{h}(Z)K(Z,Z^\dagger)]+\mathbb{E}|K(Z,Z^\dagger)|\mathbb{E}[h(Z)]\mathbb{E}U, \\
&\quad \mathrm{ICM}(U\mid Z)-\mathbb{E}[\widetilde{U}^\dagger\widetilde{h}(Z)K(Z,Z^\dagger)]+\mathbb{E}|K(Z,Z^\dagger)|\big((\mathbb{E}U)^2-\mathbb{E}[h(Z)]\mathbb{E}U\big)\Big]^\top \\
&:= [\delta_h^{*(1)},\ \delta_h^{*(2)}]^\top.
\end{aligned}$$

Thus,

$$\delta_h^{*(1)}+\delta_h^{*(2)} = \mathrm{ICM}(U\mid Z)+(\mathbb{E}U)^2\mathbb{E}|K(Z,Z^\dagger)|.$$

If $\mathbb{E}[U|Z]=0$ *a.s.*, then $\mathbb{E}U=0$ by the LIE. This implies that $\delta_h^*=0$. Conversely, if $\delta_h^*=0$, then we must have $\delta_h^{*(1)}+\delta_h^{*(2)}=0$. Note that both $\mathrm{ICM}(U\mid Z)$ and $(\mathbb{E}U)^2\mathbb{E}|K(Z,Z^\dagger)|$ are non-negative, thus we must have $\mathrm{ICM}(U\mid Z)=0$ and $(\mathbb{E}U)^2\mathbb{E}|K(Z,Z^\dagger)|=0$, corresponding to $\mathbb{E}[U|Z]=\mathbb{E}U$ *a.s.* (by the omnibus property (2)) and $\mathbb{E}U=0$, respectively. Hence $\mathbb{E}[U|Z]=0$. *a.s.*

■

S.2.2 Proof of Lemma 3.1

The estimand δ_h can be expressed as

$$\delta_h = \begin{bmatrix} \delta_h^{(1)} \\ \mathrm{ICM}(U \mid Z) - \delta_h^{(1)} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \delta_h^{(1)} \\ \mathrm{ICM}(U \mid Z) \end{bmatrix}$$

where $\delta_h^{(1)} := \mathbb{E}\big[K(Z, Z^\dagger)(h(Z^\dagger) - \mathbb{E}[h(Z)])(U - \mathbb{E}U)\big]$.

Estimation of δ_h follows by applying the analog and plug-in principles:

$$\widehat{\delta}_h = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \widehat{\delta}_h^{(1)} \\ \mathrm{ICM}_n(U \mid Z) \end{bmatrix}$$

where

$$\begin{aligned} \widehat{\delta}_h^{(1)} &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} K(Z_i, Z_j)(h(Z_j) - \mathbb{E}_n[h(Z)])U(X_i; \widehat{\theta}_n) \text{ and} \\ \mathrm{ICM}_n(U \mid Z) &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} K(Z_i, Z_j)U(X_i; \widehat{\theta}_n)U(X_j; \widehat{\theta}_n). \end{aligned}$$

Thanks to Assumption 3.3, Assumption 3.4, and the continuous mapping theorem,

$$\begin{aligned} \widehat{\delta}_h^{(1)} &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} K(Z_i, Z_j)\widetilde{h}(Z_j)\widetilde{U}_i - \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} K(Z_i, Z_j)\widetilde{h}(Z_j)G(X_i; \bar{\theta}_n)^\top(\widehat{\theta}_n - \theta_o) \\ &- \Big(\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} K(Z_i, Z_j)\widetilde{U}_i\Big)\mathbb{E}_n[\widetilde{h}(Z)] + \Big(\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} K(Z_i, Z_j)G(X_i; \bar{\theta}_n)\Big)^\top(\widehat{\theta}_n - \theta_o)\mathbb{E}_n[\widetilde{h}(Z)] \\ &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} K(Z_i, Z_j)\big(\widetilde{h}(Z_j)\widetilde{U}_i + \widetilde{h}(Z_i)\widetilde{U}_j\big)/2 \\ &- \begin{bmatrix} \mathbb{E}\big[K(Z, Z^\dagger)\widetilde{h}(Z^\dagger)G(X; \theta_o)\big] \\ \mathbb{E}\big[K(Z, Z^\dagger)\widetilde{U}\big] \end{bmatrix}^\top \begin{bmatrix} \widehat{\theta}_n - \theta_o \\ \mathbb{E}_n[h(Z)] - \mathbb{E}[h(Z)] \end{bmatrix} + o_p(n^{-1/2}) \\ &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \psi_h(D_i, D_j) - 2H_1^\top\big((\widehat{\theta}_n - \theta_o)^\top, (\mathbb{E}_n[h(Z)] - \mathbb{E}[h(Z)])\big)^\top + o_p(n^{-1/2}) \end{aligned}$$

where $\bar{\theta}_n$ satisfies $||\bar{\theta}_n - \theta_o|| \leq ||\widehat{\theta}_n - \theta_o||$ and $H_1 := (1/2)\begin{bmatrix}\mathbb{E}\big[K(Z, Z^\dagger)\widetilde{h}(Z^\dagger)G(X;\theta_o)\big] \\ \mathbb{E}\big[K(Z, Z^\dagger)\widetilde{U}\big]\end{bmatrix}$.

The first summand in the above decomposition is a second-order U-statistic. Recall $\psi_h^{(1)}(D_i) = \mathbb{E}[\psi_h(D_i, D_j)|D_i]$, then by Assumption 3.1, Assumption 3.2, and the Hoeffding decomposition (e.g., Theorem 3 in section 1.3 of Lee (1990)),

$$\begin{aligned}
&\frac{1}{n(n-1)}\sum_{i\neq j}^{n}\psi_h(D_i, D_j) = \binom{n}{2}^{-1}\sum_{i<j}^{n}\psi_h(D_i, D_j) \\
=&\delta_h^{(1)} + \frac{2}{n}\sum_{i=1}^{n}[\psi_h^{(1)}(D_i) - \delta_h^{(1)}] + \binom{n}{2}^{-1}\sum_{i<j}^{n}[\psi_h(D_i, D_j) - \psi_h^{(1)}(D_i) - \psi_h^{(1)}(D_j) + \delta_h^{(1)}] \\
=&\delta_h^{(1)} + \frac{2}{n}\sum_{i=1}^{n}[\psi_h^{(1)}(D_i) - \delta_h^{(1)}] + o_p(n^{-1/2}).
\end{aligned}$$

In addition to Assumption 3.4, $\sqrt{n}\big(\widehat{\delta}_h^{(1)} - \delta_h^{(1)}\big) = \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\big[\psi_h^{(1)}(D_i) - \delta_h^{(1)} - H_1^\top[\xi_{\theta,i}^\top, \widetilde{h}(Z_i)]^\top\big] + o_p(1)$.

Under the conditions of Lemma S.1.1,

$$\sqrt{n}[\mathrm{ICM}_n(U \mid Z) - \mathrm{ICM}(U \mid Z)] = \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\Big\{\psi_U^{(1)}(D_i) - \mathrm{ICM}(U \mid Z) - H_2^\top\xi_{\theta,i}\Big\} + o_p(1).$$

Thus, as claimed,

$$\sqrt{n}\big(\widehat{\delta}_h - \delta_h\big) = \begin{bmatrix}1 & 0 \\ -1 & 1\end{bmatrix}\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\begin{bmatrix}\psi_h^{(1)}(D_i) - \delta_h^{(1)} - H_1^\top[\xi_{\theta,i}^\top, \widetilde{h}(Z_i)]^\top \\ \psi_U^{(1)}(D_i) - \mathrm{ICM}(U \mid Z) - H_2^\top\xi_{\theta,i}\end{bmatrix} + o_p(1).$$

■

S.2.3 Proof of Theorem 3.1

Under the conditions of Lemma 3.1,

$$\sqrt{n}(\widehat{\delta}_h - \delta_h) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_h(D_i) + o_p(1) = \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\begin{bmatrix}\xi_{h,1}(D_i) \\ \xi_{h,2}(D_i) - \xi_{h,1}(D_i)\end{bmatrix} + o_p(1).$$

Thanks to the LIE, we have the following:

$$\begin{aligned}
\psi_h^{(1)}(D_i) &= \frac{1}{2}\mathbb{E}[K(Z_i,Z_j)\widetilde{h}(Z_j)|Z_i]\widetilde{U}_i + \frac{1}{2}\mathbb{E}\big[K(Z_i,Z_j)\mathbb{E}[\widetilde{U}_j \mid Z_j] \mid Z_i\big]\widetilde{h}(Z_i);\\
\psi_U^{(1)}(D_i) &= \mathbb{E}\big[K(Z_i,Z_j)\mathbb{E}[\widetilde{U}_j \mid Z_j]|Z_i\big]\widetilde{U}_i;\\
\mathrm{ICM}(U \mid Z) &= \mathbb{E}\big[K(Z,Z^\dagger)\mathbb{E}[\widetilde{U}^\dagger \mid Z^\dagger]\mathbb{E}[\widetilde{U} \mid Z]\big];\ \text{and}\\
H_2 &= \mathbb{E}\big[K(Z,Z^\dagger)\mathbb{E}[\widetilde{U}^\dagger \mid Z^\dagger]G(X;\theta_o)\big].
\end{aligned} \tag{S.22}$$

(i) Under $\mathbb{H}_o$, $\mathbb{E}[\widetilde{U} \mid Z] = 0$ *a.s.* and we have $\delta_h = 0$. From (S.22), $\xi_{h,2}(D_i) := \psi_U^{(1)}(D_i) - \mathrm{ICM}(U \mid Z) - H_2^\top \xi_{\theta,i} = 0$ *a.s.* as each summand of $\xi_{h,2}(D_i)$ is zero *a.s.* under $\mathbb{H}_o$. This implies that under $\mathbb{H}_o$,

$$\sqrt{n}\widehat{\delta}_h = [1,\ -1]^\top \frac{2}{\sqrt{n}}\sum_{i=1}^n \xi_{h,1}(D_i) + o_p(1).$$

Asymptotic normality thus follows from Assumption 3.1, Assumption 3.2, Assumption 3.4, and the Lindberg-Lévy Central Limit Theorem.

(ii) Under $\mathbb{H}_{an}$, $\sqrt{n}\delta_h^{(1)} = \mathbb{E}\big[\widetilde{h}(Z)a(Z^\dagger)K(Z,Z^\dagger)\big]$ and $\sqrt{n}\mathrm{ICM}(U \mid Z) = \mathbb{E}\big[a(Z)a(Z^\dagger)K(Z,Z^\dagger)\big]/\sqrt{n}$. Thus, under Assumption 3.2, the $\mathrm{ICM}(U \mid Z)$ part does not contribute to the local power since $\lim\limits_{n\to\infty}\sqrt{n}\mathrm{ICM}(U \mid Z) = 0$ under $\mathbb{H}_{an}$. Thus, as $n \to \infty$ under $\mathbb{H}_{an}$,

$$\begin{aligned}
\sqrt{n}\delta_h &= \sqrt{n}\begin{bmatrix}\delta_h^{(1)}\\ \mathrm{ICM}(U \mid Z) - \delta_h^{(1)}\end{bmatrix} = \begin{bmatrix}1 & 0\\ -1 & 1\end{bmatrix}\begin{bmatrix}\sqrt{n}\delta_h^{(1)}\\ \sqrt{n}\mathrm{ICM}(U \mid Z)\end{bmatrix} = \begin{bmatrix}1 & 0\\ -1 & 1\end{bmatrix}\begin{bmatrix}\mathbb{E}\big[K(Z,Z^\dagger)h(Z^\dagger)a(Z)\big]\\ \mathbb{E}\big[a(Z)a(Z^\dagger)K(Z,Z^\dagger)\big]/\sqrt{n}\end{bmatrix}\\
&\to \mathbb{E}\big[K(Z,Z^\dagger)h(Z^\dagger)a(Z)\big]\begin{bmatrix}1\\ -1\end{bmatrix} := a_o.
\end{aligned}$$

Furthermore, by the LIE and Assumption 3.4

$$\begin{aligned}
\mathbb{E}[\xi_{h,1}(D_i)] &= \mathbb{E}\big[\psi_h^{(1)}(D_i) - \delta_h^{(1)}\big] - H_1^\top \mathbb{E}[[\xi_{\theta,i}^\top, \widetilde{h}(Z_i)]^\top] = 0 \text{ and}\\
\mathbb{E}[\xi_{h,2}(D_i)] &= \mathbb{E}\big[\psi_U^{(1)}(D_i) - \mathrm{ICM}(U \mid Z)\big] - H_2^\top \mathbb{E}[\xi_{\theta,i}] = 0,
\end{aligned}$$

i.e., $\mathbb{E}[\xi_h(D_i)] = 2\mathbb{E}\big[\xi_{h,1}(D_i),\ \xi_{h,2}(D_i) - \xi_{h,1}(D_i)\big]^\top = 0$.

From the foregoing, we obtain that

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

Under $\mathbb{H}_{an}$, Assumption 3.1, and Assumption 3.2, asymptotic normality follows. Since $\mathbb{E}[\widetilde{U}|Z] = n^{-1/2}a(Z)$ under $\mathbb{H}_{an}$ it follows from (S.22) that

$$\mathrm{Var}(\xi_h(D_i)) - \Omega_{h,o} = O(n^{-1/2}),$$

and the continuous mapping theorem applies.

(iii) Under the local alternative $\mathbb{H}'_{an} : \mathbb{E}[\widetilde{U}|Z] = n^{-1/4}a(Z)$, we have $\sqrt{n}\mathrm{ICM}(U \mid Z) = \mathbb{E}\big[a(Z)a(Z^\dagger)K(Z,Z^\dagger)\big] > 0$ and $\sqrt{n}\delta_h^{(1)} = n^{1/4}\mathbb{E}\big[\widetilde{h}(Z)a(Z^\dagger)K(Z,Z^\dagger)\big] = 0$ if $\mathbb{E}\big[a(Z^\dagger)K(Z,Z^\dagger) \mid Z\big]$ is orthogonal to $h(Z)$. By arguments akin to part (ii) above for $\mathbb{E}[\widetilde{U}|Z] = n^{-1/4}a(Z)$ under $\mathbb{H}'_{an}$, it follows from (S.22) that

$$\mathrm{Var}(\xi_h(D_i)) - \Omega_{h,o} = O(n^{-1/4}).$$

Invoking the continuous mapping theorem completes this part of the proof.

(iv) $\delta_h \neq 0$ under $\mathbb{H}_a$. Thanks to the asymptotically linear expression in Lemma 3.1, the result follows from Assumptions 3.1 to 3.4.

■

S.2.4 Proof of Theorem 3.2

By Theorem 2, equation (3.3) in Maesono (1998), we know that

$$\widetilde{\Omega}_{h,n} - \Omega_h = O_p(n^{-1/2}),$$

for both $\mathbb{H}_o$ and $\mathbb{H}_a$. Furthermore, under $\mathbb{H}_{an}$, we know that $\Omega_h - \Omega_{h,o} = O(n^{-1/2})$. Hence, for any small $\iota \in (0, 1/2)$, $n^{1/2-\iota}(\widetilde{\Omega}_{h,n} - \Omega_{h,o}) = o_p(1)$, under $\mathbb{H}_o$ and $\mathbb{H}_{an}$. This implies that Assumption 2.2 in Dufour and Valéry (2016) holds by setting $b_n = n^{1/2}$ and $c_n = Cn^{-1/2+\iota}$ for some constant $C > 0$. Furthermore, by Proposition 9.1 in Dufour and Valéry (2016), we know that $\widehat{\Omega}^{-}_{h,n} \xrightarrow{p} \Omega^{-}_{h,o}$ under $\mathbb{H}_o$ and $\mathbb{H}_{an}$. Similarly, $\widehat{\Omega}^{-}_{h,n} \xrightarrow{p} \Omega^{-}_{h,a}$ under $\mathbb{H}_a$.

(i). By Corollary 9.3 in Dufour and Valéry (2016), $T_{h,n} \xrightarrow{d} \chi^2_1$.

(ii). By the continuous mapping theorem and Theorem 3.1 (ii),

$$T_{h,n} \xrightarrow{d} \chi^2_1(b_o).$$

(iii) By the continuous mapping theorem and Theorem 3.1 (iii),

$$T_{h,n} \xrightarrow{d} \chi^2_1(b'_o).$$

(iv). When $\delta_h \notin \mathcal{M}_0$, we have that $\lim\limits_{n\to\infty} T_{h,n} = \lim\limits_{n\to\infty} n\widehat{\delta}_h^\top \Omega_{h,a}^- \widehat{\delta}_h = \infty$, in probability, and the result follows.

■

S.2.5 Proof of the Result in Remark 3.3

Lemma S.2.1. *$\delta_h \notin \mathcal{M}_0$ for V_h in (6).*

Proof. Under $\mathbb{H}_a$, $a(Z) = \mathbb{E}[U|Z] - \mathbb{E}[U]$ is non-degenerate with $\mathbb{E}a(Z) = 0$. Recall $\widetilde{h}(Z) := h(Z) - \mathbb{E}[h(Z)]$, and define

$$
\begin{aligned}
m_{\widetilde{V}}(Z) &:= \Big[\mathbb{E}\{K(Z, Z^\dagger)\widetilde{h}(Z^\dagger)|Z\},\ \mathbb{E}\{K(Z, Z^\dagger)[a(Z^\dagger) - \widetilde{h}(Z^\dagger)]|Z\}\Big]^\top \text{ and} \\
m_{\widetilde{U}}(Z) &:= \mathbb{E}\big[K(Z, Z^\dagger)a(Z^\dagger)|Z\big].
\end{aligned}
$$

First, when Ω_a is of rank 2, the claim holds because $\delta_h \neq \mathbf{0}^\top$ under $\mathbb{H}_a$.

Second, recall that $\Omega_a = \mathrm{Var}\{\xi_h(D)\}$ using the expression $\xi_h(D) = [m_{\widetilde{U}}(Z) - \mathbb{E}m_{\widetilde{U}}(Z)](V_h - \mathbb{E}V_h) + [m_{\widetilde{V}}(Z) - \mathbb{E}m_{\widetilde{V}}(Z)](U - \mathbb{E}U)$ and $\Omega_h = \mathrm{Var}\left\{\xi_h(D)\right\}$. By the Cauchy-Schwarz inequality, we know that Ω_a is of rank one if and only if for some constant $C \neq -1$,

$$
\xi_h(D)^\top[-C, 1] = 0 \ a.s.
$$

Here we rule out $C = -1$ as it indicates that $\mathrm{ICM}(a(Z)|Z) = 0$, a contradiction of $\mathbb{H}_a$.

Let $f(Z^\dagger) := a(Z^\dagger) - (C+1)\widetilde{h}(Z^\dagger)$, this implies that

$$
[m_{\widetilde{U}}(Z) - \mathbb{E}m_{\widetilde{U}}(Z)][\widetilde{U} - (C+1)\widetilde{h}(Z)] + [m_{\widetilde{f}}(Z) - \mathbb{E}m_{\widetilde{f}}(Z)]\widetilde{U} = 0 \ a.s. \tag{S.23}
$$

where $m_{\widetilde{f}}(Z) = \mathbb{E}\{K(Z, Z^\dagger)f(Z^\dagger)|Z\}$.

Recall that $\mathbb{E}[\widetilde{U}|Z] = a(Z)$, by taking conditional expectation of (S.23) w.r.t. Z, we have

$$
[m_{\widetilde{U}}(Z) - \mathbb{E}m_{\widetilde{U}}(Z)]f(Z) + [m_{\widetilde{f}}(Z) - \mathbb{E}m_{\widetilde{f}}(Z)]a(Z) = 0 \ a.s. \tag{S.24}
$$

Therefore, by taking the difference, we have

$$
\Big\{m_{\widetilde{U}}(Z) + m_{\widetilde{f}}(Z) - \mathbb{E}m_{\widetilde{U}}(Z) + \mathbb{E}m_{\widetilde{f}}(Z)\Big\}\,[\widetilde{U} - a(Z)] = 0 \ a.s.
$$

In light of the foregoing, either one of the following holds:

(1) $\widetilde{U} = a(Z)$ *a.s.*

(2) $\widetilde{U} \neq a(Z)$, so that the coefficient on $\widetilde{U}$ in (S.23) is zero, i.e.

$$\left\{m_{\widetilde{U}}(Z) + m_{\widetilde{f}}(Z) - \mathbb{E}m_{\widetilde{U}}(Z) + \mathbb{E}m_{\widetilde{f}}(Z)\right\} = 0 \ a.s.$$

which further implies that $[m_{\widetilde{U}}(Z) - \mathbb{E}m_{\widetilde{U}}(Z)]\widetilde{h}(Z) = 0$ *a.s.* in combination of (S.23).

We proceed using proof by contradiction.

If (1) holds, we have

$$\Omega_a = \operatorname{Var}\left\{[m_{\widetilde{U}}(Z) - \mathbb{E}m_{\widetilde{U}}(Z)]\widetilde{h}(Z) + [m_{\widetilde{h}}(Z) - \mathbb{E}m_{\widetilde{h}}(Z)]a(Z)|D]\right\}\begin{bmatrix}1 & C\\ C & C^2\end{bmatrix}$$

which implies $\mathcal{M}_0 = \{(x,y)^\top | x = -Cy, y \in \mathbb{R}\}$. Note in this case, by taking the expectation of (S.24), we can show that $\mathbb{E}\{K(Z,Z^\dagger)a(Z^\dagger)[a(Z) - \widetilde{h}(Z)]\} = C\mathbb{E}[K(Z,Z^\dagger)a(Z^\dagger)\widetilde{h}(Z)]$, so that $\delta_h = \mathbb{E}[K(Z,Z^\dagger)a(Z^\dagger)\widetilde{h}(Z)][1,C]^\top \notin \mathcal{M}_0$.

If (2) holds, then we have $m_{\widetilde{U}}(Z) - \mathbb{E}m_{\widetilde{U}}(Z) = 0$ *a.s.* since $\widetilde{h}(Z)$ is non-degenerate. This implies that $\mathbb{E}\{[m_{\widetilde{U}}(Z) - \mathbb{E}m_{\widetilde{U}}(Z)]a(Z)\} = \mathrm{ICM}(U|Z) = 0$, a contradiction.

■

S.2.6 Proof of Theorem 3.3

For two functions $f_a(x)$ and $f_b(x)$, write $f_a(x) \sim f_b(x)$ if and only if $f_a(x)/f_b(x) \to 1$ as $x \to \infty$. By Zolotarev (1961), we know that

$$\log \mathbb{P}\Big(\sum_{k=1}^{\infty} \lambda_k G_k^2 > x\Big) \sim -x/(2\lambda_1), \quad \text{as } x \to \infty.$$

Clearly, under $\mathbb{H}_a$, we have $n\mathrm{ICM}_n(U \mid Z) \to \infty$ in probability. Thus,

$$c_G = \operatorname*{plim}_{n\to\infty} \frac{n\mathrm{ICM}_n(U \mid Z)}{n\lambda_1} = \frac{\mathrm{ICM}(U \mid Z)}{\lambda_1},$$

where we note that $\mathrm{ICM}_n(U \mid Z) \to_{a.s.} \mathrm{ICM}(U \mid Z)$ under Assumption 3.2.

Next, by the large deviation result for the chi-squared distribution,

$$\log \mathbb{P}\big(\chi_1^2 > x\big) \sim -x/2 \quad \text{as } x \to \infty.$$

Also, $\operatorname*{plim}_{n\to\infty} T_{h,n} = \operatorname*{plim}_{n\to\infty} n\widehat{\delta}_h^\top \widehat{\Omega}_{h,n}^- \widehat{\delta}_h = \lim_{n\to\infty} n\delta_a^\top \Omega_{h,a}^- \delta_a = \infty$ under $\mathbb{H}_a$, thus

$$c_T = \lim_{n\to\infty} \frac{2}{n} \frac{n\delta_a^\top \Omega_{h,a}^- \delta_a}{2} = \delta_a^\top \Omega_{h,a}^- \delta_a.$$

■

S.3 Asymptotic properties of $\widehat{\delta}_h^*$.

A natural empirical estimator for δ_h^* is given by $\widehat{\delta}_h^* = \widehat{\delta}_h + \widehat{\eta}_h^*$ where

$$\widehat{\eta}_h^* := \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j\neq i} |K(Z_i, Z_j)| \mathbb{E}_n[U] \big[\mathbb{E}_n[h(Z)],\ \mathbb{E}_n[U] - \mathbb{E}_n[h(Z)]\big]^\top.$$

Accordingly, define $\eta_h^* := \mathbb{E}\big[|K(Z, Z^\dagger)|\big]\mathbb{E}U\mathbb{E}\big[h(Z),\ U - h(Z)\big]^\top$. We consider the simpler case of mean independence with nullity – specification with nullity testing follows straightforwardly. For the asymptotic analysis of $\widehat{\delta}_h^*$, we consider the following sequence of local alternatives:

$$\mathbb{H}_{an}^* : \mathbb{E}[U \mid Z] = n^{-1/2} a^*(Z),$$

where $a^*(Z)$ is a nonzero, possibly degenerate function of Z.

Thanks to the Hoeffding decomposition

$$\begin{aligned}
\sqrt{n}(\widehat{\eta}_h^* - \eta_h^*) &= \frac{1}{\sqrt{n}(n-1)} \sum_{i\neq j} \begin{bmatrix} \mathbb{E}[U]\mathbb{E}[h(Z)] \\ (\mathbb{E}[U])^2 - \mathbb{E}[U]\mathbb{E}[h(Z)] \end{bmatrix} \big(|K(Z_i, Z_j)| - \mathbb{E}\big[|K(Z, Z^\dagger)|\big]\big) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{bmatrix} \mathbb{E}[h(Z)] \\ \mathbb{E}[U] - \mathbb{E}[h(Z)] \end{bmatrix} \mathbb{E}\big[|K(Z, Z^\dagger)|\big] \big(U_i - \mathbb{E}[U]\big) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{E}\big[|K(Z, Z^\dagger)|\big] \mathbb{E}[U] \begin{bmatrix} h(Z_i) - \mathbb{E}[h(Z_i)] \\ \big(U_i - \mathbb{E}[U]\big) - \big(h(Z_i) - \mathbb{E}[h(Z)]\big) \end{bmatrix} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Bigg\{ \begin{bmatrix} \mathbb{E}[U]\mathbb{E}[h(Z)] \\ (\mathbb{E}[U])^2 - \mathbb{E}[U]\mathbb{E}[h(Z)] \end{bmatrix} 2\big(\mathbb{E}\big[(|K(Z_i, Z_j)|) \mid Z_i\big] - \mathbb{E}\big[|K(Z, Z^\dagger)|\big]\big) \\
&\qquad + \begin{bmatrix} \mathbb{E}[h(Z)] \\ \mathbb{E}[U] - \mathbb{E}[h(Z)] \end{bmatrix} \mathbb{E}\big[|K(Z, Z^\dagger)|\big] \big(U_i - \mathbb{E}[U]\big)
\end{aligned}$$

$$
\left. + \mathbb{E}\big[|K(Z,Z^{\dagger})|\big]\mathbb{E}[U]\begin{bmatrix} h(Z_i)-\mathbb{E}[h(Z)] \\ \big(U_i-\mathbb{E}[U]\big)-\big(h(Z_i)-\mathbb{E}[h(Z)]\big)\end{bmatrix}\right\} + o_p(1)
$$

$$
=: \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_h^*(D_i)+o_p(1).
$$

In addition to Lemma 3.1, the following is the asymptotically linear representation of $\widehat{\delta}_h^*$:

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

As $n\to\infty$ under $\mathbb{H}_{an}^*$,

$$
\sqrt{n}\delta_h^*=\sqrt{n}\delta_h+\sqrt{n}\eta_h^*\to\begin{bmatrix}1\\-1\end{bmatrix}\Big(\mathbb{E}\big[h(Z)\big(a^*(Z^{\dagger})-\mathbb{E}[a^*(Z^{\dagger})]\big)K(Z,Z^{\dagger})\big]+\mathbb{E}\big[|K(Z,Z^{\dagger})|\big]\mathbb{E}[a^*(Z)]\mathbb{E}[h(Z)]\Big):=a_o^*
$$

where $\lim_{n\to\infty}\sqrt{n}\eta_h^*=\mathbb{E}\big[|K(Z,Z^{\dagger})|\big]\mathbb{E}[a^*(Z)]\mathbb{E}[h(Z)]\big[1,-1\big]^{\top}$.

Theorem S.3.1. *Let the conditions of Theorem 3.1 hold, then*

(i) under $\mathbb{H}_o^*$, $\sqrt{n}\widehat{\delta}_h^*\xrightarrow{d}\mathcal{N}(0,\Omega_{h,o}^*)$;

(ii) under $\mathbb{H}_{an}^*$, $\sqrt{n}\widehat{\delta}_h^*\xrightarrow{d}\mathcal{N}(a_o^*,\Omega_{h,o}^*)$; *and*

(iii) under $\mathbb{H}_a^*$; $\sqrt{n}(\widehat{\delta}_h^*-\delta_h^*)\xrightarrow{d}\mathcal{N}(0,\Omega_{h,a}^*)$,

where $\Omega_{h,o}^*$ *and* $\Omega_{h,a}^*$ *correspond to specific expressions of* $\Omega_h^*=\operatorname{Var}\big[\xi_h(D)+\phi_h^*(D)\big]$ *under* $\mathbb{H}_o^*$ *and* $\mathbb{H}_a^*$, *respectively.*

Proof. The proof follows that of Theorem 3.1 and the discussion preceding the theorem. The details are therefore omitted.

■

S.4 Numerical Illustration and Computation of the Bahadur slopes

This section presents additional details concerning the Bahadur slopes of the χ^2-test and the bootstrap-based ICM specification tests. The first subsection offers concrete examples that illustrate Theorem 3.3, while the second subsection describes the numerical integration procedure used in the first subsection.

S.4.1 Numerical Illustration

To make the Theorem 3.3 results concrete, we consider the simple design

$$U = \exp(-Z^2/3) - \sqrt{3/5} + \mathcal{E}, \quad \text{with } Z \sim \mathcal{N}(0,1) \quad \text{and } \mathcal{E} \sim \mathcal{U}[-\sqrt{3}, \sqrt{3}],$$

where $\mathcal{E}$ is independent of Z, hence $\mathbb{E}[U \mid Z] := a(Z) = \exp(-Z^2/3) - \sqrt{3/5}$, and $\mathbb{E}U = 0$. To examine the power behavior of the competing tests, we introduce variations in the construction of V. Using the Gaussian kernel $K(Z, Z^\dagger) = \exp\big(-0.5(Z - Z^\dagger)^2\big)$ for both tests, Table S.1 shows the Bahadur slopes of the bootstrap-based ICM test alongside the χ^2-test with

(1) $V_1 = [h_1(Z),\ U - h_1(Z)]^\top$, $h_1(Z) = \exp(Z)$ which is agnostic of $a(Z)$;

(2) $V_{1a} = [h_1(Z),\ U - h_1(Z),\ Z]^\top$;

(3) $V_2 = [h_2(Z),\ U - h_2(Z)]^\top$, $h_2(Z) = a(Z)$ which results in a singular covariance matrix under $\mathbb{H}_a$;

(4) $V_3 = [h_3(Z),\ U - h_3(Z)]^\top$, $h_3(Z) = \sqrt{3}\exp(-Z^2/2)$ which satisfies $\mathbb{E}[K(Z, Z^\dagger)h(Z^\dagger) \mid Z] = \exp(-Z^2/3) - \sqrt{3/5} := a(Z)$; and

(5) $V_{3a} = [h_3(Z),\ U - h_3(Z),\ Z]^\top$.

In scenarios (3) - (5), we use prior knowledge of the alternative. In (3), it can be shown that $a_o = [\mathrm{ICM}(U \mid Z), 0]^\top$, and

$$\Omega_{h,a} = A \times \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix},$$

where $A = \mathrm{Var}\Big\{\widetilde{\xi}_{h,2}(D)\Big\}$ and $\widetilde{\xi}_{h,2}(D) = (U - \mathbb{E}U)\mathbb{E}[a(Z^\dagger)K(Z, Z^\dagger) \mid Z]$. Clearly, $\mathbb{E}[\widetilde{\xi}_{h,2}(D)] = \mathrm{ICM}(U \mid Z)$, therefore $a_o^\top \Omega_{h,a}^- a_o = \mathrm{ICM}^2(U \mid Z)/[4\,\mathrm{Var}(\xi_{h,2}(D))]$. This is in fact a worst-case scenario under $\mathbb{H}_a$ for the χ^2-test. To fully make use of the information of $a(Z)$, the choice of $h(Z)$ in scenario (4) maximizes the linear dependence with $\mathbb{E}\big[K(Z, Z^\dagger)(U^\dagger - \mathbb{E}U) \mid Z\big]$ in the first element of δ_h while the second term is not degenerate: $\delta_h = \mathbb{E}\big[a^2(Z),\ \mathrm{ICM}(U \mid Z) - a^2(Z)\big]^\top$. Therefore, power should be augmented. Scenarios (2) and (5) are Vs of scenarios (1) and (4) augmented with Z.

Table S.1 shows that the bootstrap-based test is an intermediate case between the agnostic choice with $h(Z) = \exp(Z)$ (scenario (1)) and the worst case with $h(Z) = a(Z)$ (scenario (3)). In the case where the linear dependence is maximized with respect to either element of $V_h = [h(Z),\ U - h(Z)]^\top$ without degeneracy in the other (scenario (4)), the χ^2-test Bahadur slope is larger. Moreover, we note that when augmenting V, one observes a slight increase in the Bahadur slope for scenario (2) but none

Table S.1: Bahadur Slopes

	Boot	χ^2				
		V_1	V_{1a}	V_2	V_3	V_{3a}
Bahadur Slope	0.0109	0.0214	0.0242	0.0056	0.0246	0.0246

for scenario (5). The latter case is not surprising as Z is orthogonal to the direction under alternative, namely, $\mathbb{E}[K(Z, Z^\dagger)h(Z^\dagger) \mid Z] = a(Z) = \exp(-Z^2/3) - \sqrt{3/5}$ is orthogonal to Z thus the augmentation with Z does not augment power.

S.4.2 Numerical Integration

Here, we briefly provide details pertaining to the Monte Carlo numerical integration used to obtain the Bahadur slopes in the preceding sub-section. The approach proceeds by computing $\mathbb{E}[a(Z)a(Z^\dagger)K(Z, Z^\dagger)]$, λ_1, a_o, and $\Omega_{h,a}$ on a sample of 1 000 random draws following the DGP in Section 3.4. $\lambda_1 = 0.6175762$ is computed using steps 1-3 of Seri (2022, Algorithm 1) with (i, j)'th element $\exp(-0.5(Z_i - Z_j)^2)\mathcal{E}_i\mathcal{E}_j/(n-1)$. The numerical values of the Bahadur slopes in Table S.1 are then obtained using averages of the quantities $\mathbb{E}[a(Z)a(Z^\dagger)K(Z, Z^\dagger)]$, λ_1, a_o, and $\Omega_{h,a}$ over 10 000 replications.

S.5 Monte Carlo Experiments - Mean Independence Test

This section examines the empirical size control and power performance of the test of mean independence via simulations. Section S.5.1 presents five DGPs; Section S.5.2 examines the size control and power performance of the tests of mean independence; Section S.5.3 examines the performance of the χ^2-test with V augmented to dimensions $p_v > 2$; Section S.5.4 conducts sensitivity analyses of the size and power performance of the test to variations of the tuning rule $c_n = \widetilde{\lambda}_1 n^{-1/3}$ and other selection criteria used in the literature; and Section S.5.5 examines the test of nullity $\mathbb{E}[U \mid Z] = \mathbb{E}[U] = 0$ *a.s.* via simulations.

S.5.1 Specifications

Five different DGPs with conditional heteroskedasticity are considered for the test of mean independence. The first four DGPs follow directly from LS1 through LS4 in the main text with $\theta_l = 0$, $l \geq 1$. The fifth DGP examines alternatives that are binary and non-monotone in Z.

MI 1: $U = 1 + \dfrac{\mathcal{E}}{\sqrt{1+Z_1^2}}$;

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

MI 3: $U = 1 + \gamma \sum_{l=1}^{5} \dfrac{\cos(2Z_l)}{\sqrt{2(1-\exp(-8))}} + \dfrac{\mathcal{E}}{\sqrt{1+Z_1^2}}$;

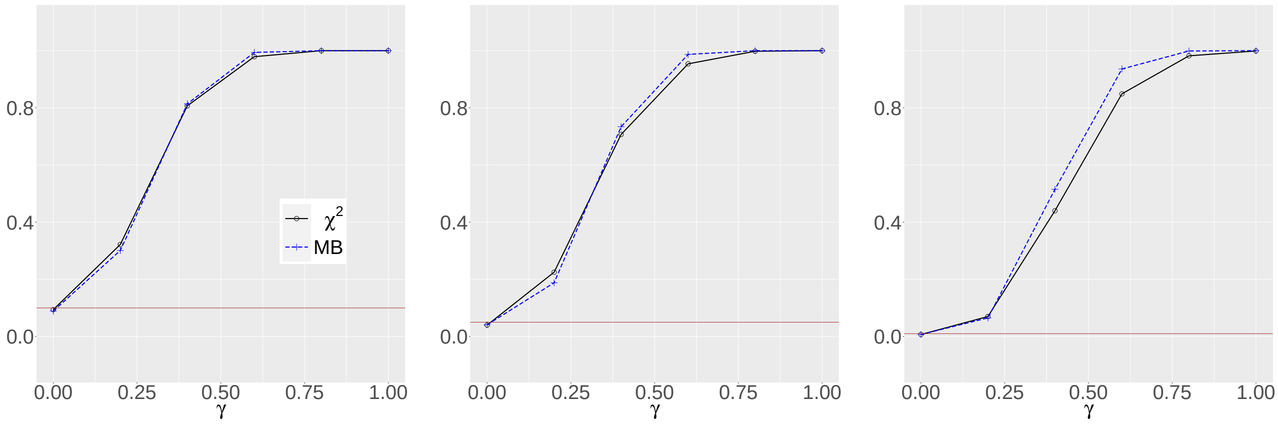
MI 4: $U = 1 + \gamma \sum_{l=1}^{5} \big(\exp(-Z_l^2/3) - \sqrt{3/5} \big) + \dfrac{\mathcal{E}}{\sqrt{1+Z_1^2}}$; and

MI 5: $U = 1 + \dfrac{\gamma}{2} \sum_{l=1}^{5} \mathrm{I}\big(\,|\, Z_l| < -\Phi^{-1}(1/4) \big) + \dfrac{\mathcal{E}}{\sqrt{1+Z_1^2}}$.

S.5.2 Empirical Size and Power

Table S.2 compares the empirical sizes of the proposed χ^2-test for mean independence with those of the multiplier bootstrap procedure, using DGP MI 1. It also reports their power under local alternatives in DGP MI 2.[1] As shown in Table S.2, both tests exhibit accurate size control and non-trivial local power across all nominal levels and ICM kernels.

Figure S.1: DGP MI 2 – Gaussian Kernel – $n = 400$.



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

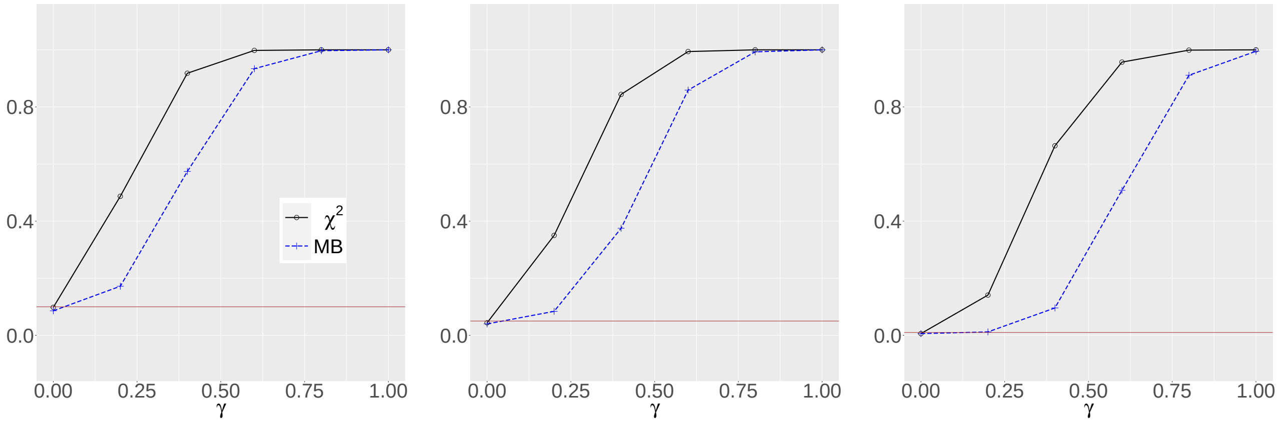
---

[1] The multiplier and wild bootstrap procedures coincide for the test of mean independence since there is no model (re)-estimation.

Table S.2: Empirical Size & Local Power

		10%		5%		1%	
n	Kernel	χ^2	MB	χ^2	MB	χ^2	MB
MI 1		Empirical Size					
200	Gauss	0.100	0.095	0.043	0.048	0.005	0.008
	Euclid	0.078	0.093	0.023	0.044	0.004	0.007
400	Gauss	0.094	0.089	0.040	0.041	0.007	0.007
	Euclid	0.098	0.086	0.044	0.04	0.006	0.006
600	Gauss	0.077	0.093	0.042	0.047	0.009	0.012
	Euclid	0.096	0.098	0.039	0.045	0.006	0.012
800	Gauss	0.095	0.089	0.047	0.041	0.004	0.005
	Euclid	0.105	0.094	0.042	0.044	0.007	0.007
MI 2		Local Power: $\gamma = 5/\sqrt{n}$					
200	Gauss	0.439	0.421	0.291	0.308	0.095	0.140
	Euclid	0.598	0.243	0.438	0.113	0.192	0.020
400	Gauss	0.464	0.435	0.318	0.306	0.125	0.111
	Euclid	0.632	0.231	0.501	0.118	0.245	0.023
600	Gauss	0.471	0.427	0.327	0.308	0.12	0.134
	Euclid	0.637	0.244	0.504	0.136	0.245	0.022
800	Gauss	0.463	0.438	0.328	0.325	0.142	0.131
	Euclid	0.628	0.233	0.504	0.130	0.259	0.024

Figure S.2: DGP MI 2 – Negative Euclidean – $n = 400$.

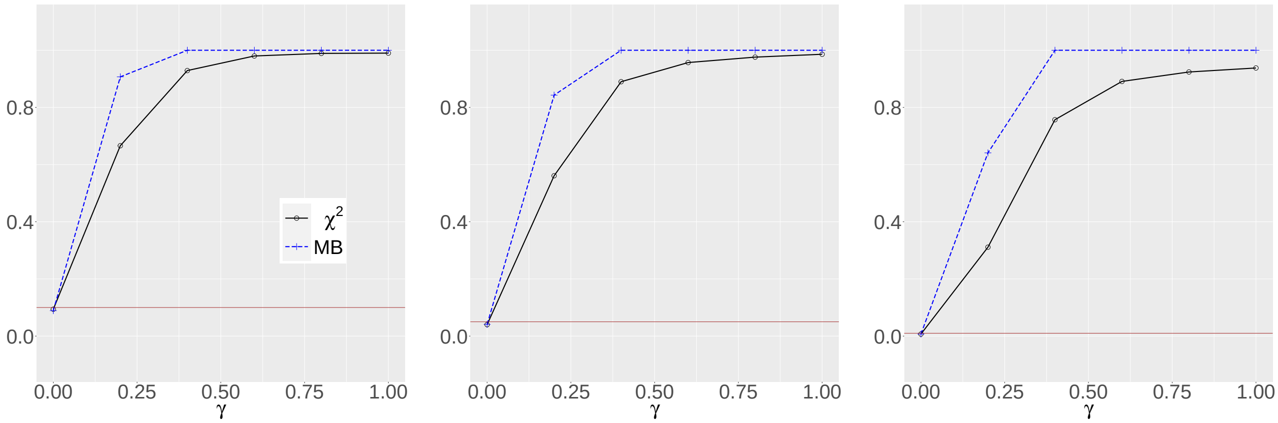


(a) 10%

(b) 5%

(c) 1%

Figure S.3: DGP MI 3 – Gaussian Kernel – $n = 400$.

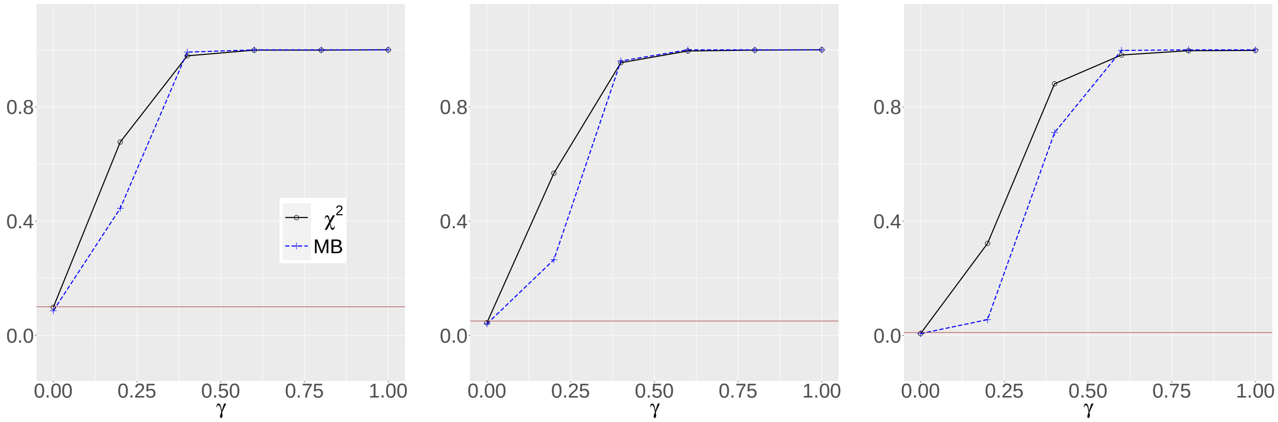


(a) 10%

(b) 5%

(c) 1%

Figure S.4: DGP MI 3 – Negative Euclidean – $n = 400$.

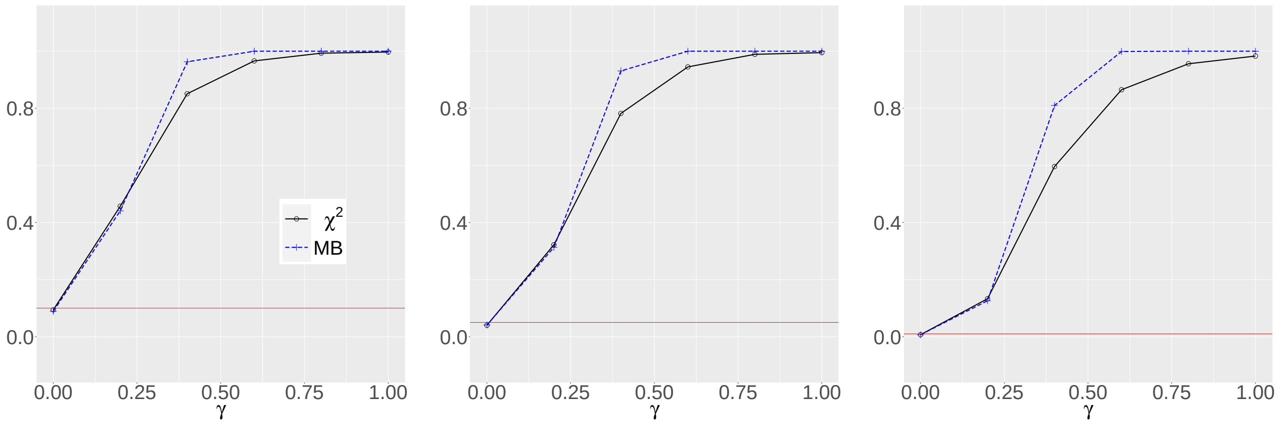


(a) 10%

(b) 5%

(c) 1%

Figure S.5: DGP MI 4 – Gaussian Kernel – $n = 400$.

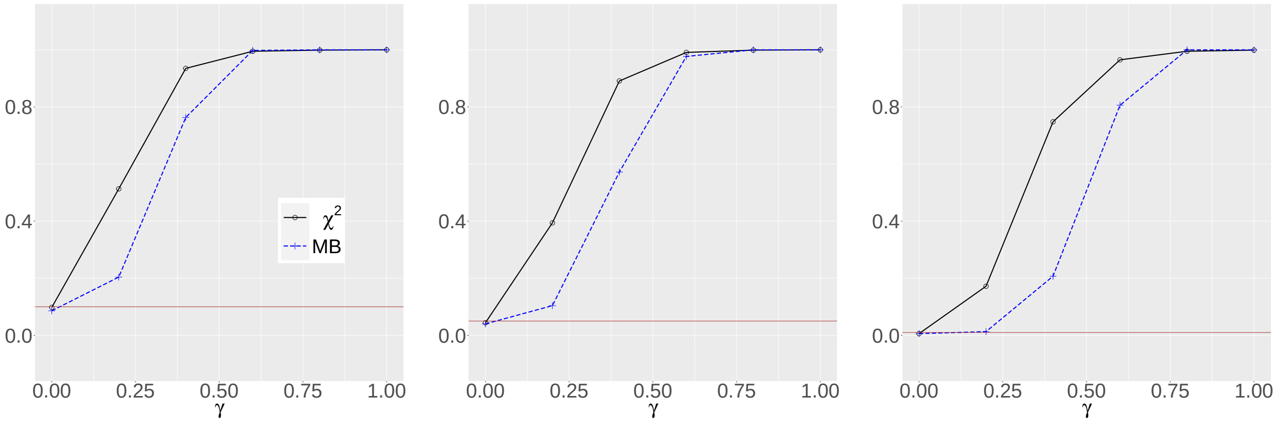


(a) 10%

(b) 5%

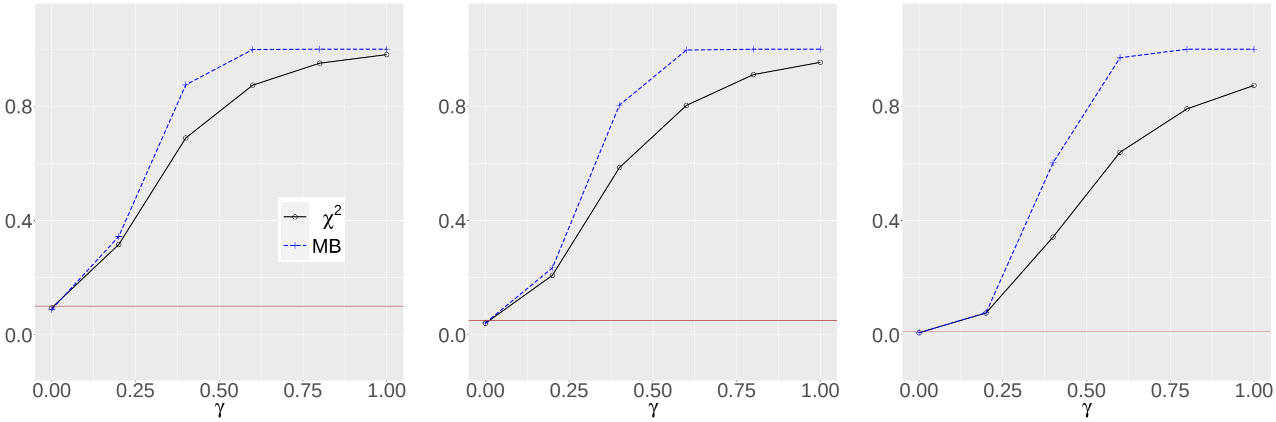
(c) 1%

Figure S.6: DGP MI 4 – Negative Euclidean – $n = 400$.



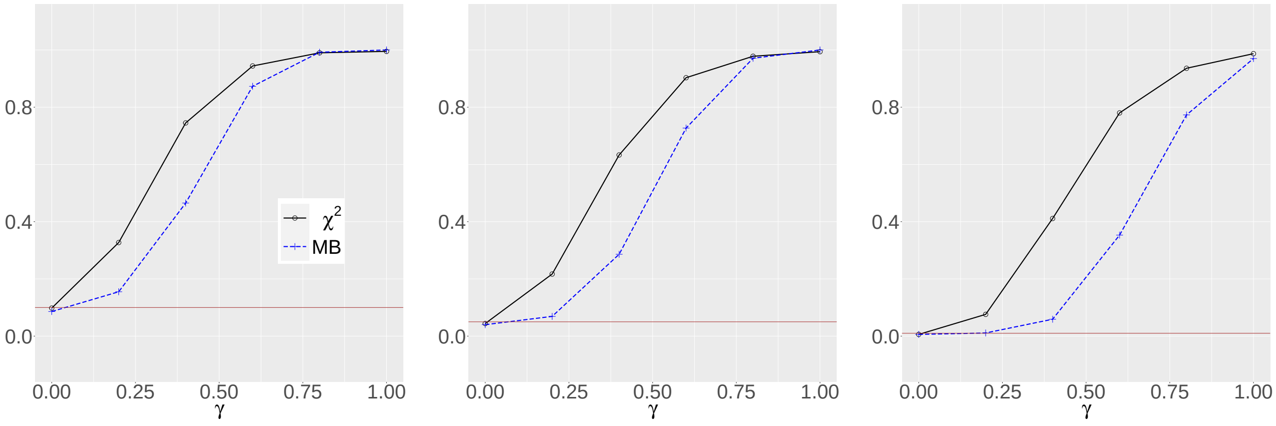
(a) 10%   (b) 5%   (c) 1%

Figure S.7: DGP MI 5 – Gaussian Kernel – $n = 400$.



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

Figure S.8: DGP MI 5 – Negative Euclidean – $n = 400$.



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

Figures S.1 to S.8 illustrate that all tests generally perform well in detecting violations of mean independence. The χ^2-test shows competitive performance across all DGPs. As indicated by Theorem 3.3, relative performance depends on, for example, the choice of kernel. Overall, no single test consistently outperforms the other, highlighting the complementary strengths of the χ^2- and bootstrap-based ICM tests.

S.5.3 $p_v > 2$

The goal of this subsection is to study the sensitivity of the χ^2-test to the dimension of V alongside bootstrap-based ICM tests. The χ^2-test is implemented in the rest of this section using the Gaussian kernel. Consider the following DGP:

MI 6: $U = 1 + \frac{\gamma}{0.233}(\exp(-Z^2/3) - \sqrt{3/5}) + \mathcal{E}$, $Z \sim \mathcal{N}(0,1)$, and $\mathcal{E} \sim \mathcal{U}[-\sqrt{3}, \sqrt{3}]$.

Define the following: $h_1(Z) := \exp(Z)$; $h_2(Z) := 4\sqrt{3}\exp(-Z^2/2)$; $V_1 := [h_1(Z),\ U - h_1(Z)]^\top$; $V_{1A} := [h_1(Z),\ U - h_1(Z),\ h_2(Z)]^\top$; $V_2 := [h_2(Z),\ U - h_2(Z)]^\top$; and $V_{2A} := [h_2(Z),\ U - h_2(Z),\ h_1(Z)]^\top$. Let $g_p(Z) \in \mathbb{R}^p$ denote a vector of orthogonal polynomials of Z with degrees 1 through p. Then, we generate higher dimensional V given by V_{1B} and V_{2B}, respectively, which augment V_{1A} and V_{2A} using $g_2(Z)$, and V_{1C} and V_{2C}, respectively, which augment V_{1A} and V_{2A} using $g_7(Z)$. When $p_v \geq 3$, the degrees of freedom of the χ^2-test is set to the number of positive eigenvalues of $\widehat{\Omega}_{h,n}$. In contrast to V_2 where $h_2(Z)$ targets the alternative, V_1 is agnostic about the alternative.

Table S.3 presents the power performance of four variations (depending on the dimension and specification of V) of the χ^2-test (based on the Gaussian kernel) in addition to bootstrap-based ICM tests

Table S.3: DGP MI 6 - Sensitivity to p_v

γ	Sig-Lev	χ^2-test								Wild Bootstrap		
		V_1	V_{1A}	V_{1B}	V_{1C}	V_2	V_{2A}	V_{2B}	V_{2C}	Gauss	MDD	Esc6
		$p_v = 2$	$p_v = 3$	$p_v = 5$	$p_v = 10$	$p_v = 2$	$p_v = 3$	$p_v = 5$	$p_v = 10$			
0.0	10%	0.112	0.112	0.102	0.100	0.113	0.116	0.110	0.107	0.100	0.100	0.104
	5%	0.051	0.051	0.050	0.050	0.058	0.058	0.056	0.056	0.049	0.053	0.053
	1%	0.007	0.007	0.007	0.007	0.008	0.008	0.009	0.007	0.009	0.009	0.007
0.2	10%	0.976	0.976	0.973	0.973	0.989	0.989	0.982	0.977	0.925	0.818	0.707
	5%	0.953	0.953	0.946	0.945	0.980	0.979	0.967	0.958	0.854	0.678	0.463
	1%	0.838	0.838	0.822	0.822	0.912	0.911	0.863	0.852	0.622	0.264	0.115
0.4	10%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
	1%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.965
0.6	10%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.8	10%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1.0	10%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

of mean independence Gauss, MDD, and Esc6. MDD and Esc6 are based on the negative Euclidean kernel of Shao and Zhang (2014) and the kernel proposed in Escanciano (2006), respectively. First, one observes good size control and non-trivial power increasing in γ for all tests. Second, increasing the dimension of V appears to decrease power. A comparison of the test with V_2 (where the alternative is targeted) and $\{V_1, V_{1A}, V_{1B}, V_{1C}\}$ shows a power advantage of targeting the alternative using a parsimonious 2-dimensional V.

S.5.4 Selection Criteria c_n

In all preceding implementations of the proposed χ^2-test, the tuning parameter in the regularized $\widehat{\Omega}_{h,n}$ is set to $c_n = \widetilde{\lambda}_1 n^{-1/3}$. This subsection conducts a robustness exercise to examine the sensitivity of the empirical size and power performance to the tuning rule c_n. Two scenarios are considered.

S.5.4.1 Scenario 1

The first scenario concerns sensitivity to the constant ι in the rule $c_n = \widetilde{\lambda}_1 n^{-\iota}$. For the implementation, the set $\iota \in \{2/5, 1/3, 1/4, 1/6\}$ is considered with $p_v = 2$, and $V = V_1$ using the following DGP:

MI 7: $U = 1 + \gamma Z^2 + \mathcal{E}$, $Z \sim \mathcal{N}(0,1)$, and $\mathcal{E} \sim \mathcal{U}[-\sqrt{3}, \sqrt{3}]$.

Table S.4 presents results that compare the performance of the χ^2-test by different choices of c_n in the first scenario. A clear conclusion is that the results are robust to the choice of ι in the rule $c_n = \widetilde{\lambda}_1 n^{-\iota}$ as there are negligible numerical differences in the empirical size and power across different valid choices of $\iota \in (0, 1/2)$.

S.5.4.2 Scenario 2

The second scenario compares the χ^2-test with, in addition to c_n in Scenario 1 above, suitable selection criteria typically used for truncated singular value decomposition – see Falini (2022) for a review. The setting adopted in this scenario is DGP MI 6, $p_v = 10$, and $V = V_{1C}$ from Section S.5.3. Let $p(c_n)$ denote the number of non-zero eigenvalues in the regularized $\widehat{\Omega}_{h,n}$. Define

$$E_l := -\frac{1}{\log(p_v)} \sum_{l'=1}^{l} \widetilde{f}_{l'} \log(\widetilde{f}_{l'}) \text{ where } \widetilde{f}_l := \widetilde{\lambda}_l^2 / \sum_{l'=1}^{p_v} \widetilde{\lambda}_{l'}^2.$$

The following suitable SVD selection criteria are defined in terms of $p(c_n)$.

Table S.4: DGP MI 7 - Sensitivity to ι

γ	Sig-Lev	$c_n = \widetilde{\lambda}_1 n^{-\iota}$			
		$\iota = \frac{2}{5}$	$\iota = \frac{1}{3}$	$\iota = \frac{1}{4}$	$\iota = \frac{1}{6}$
0.0	10%	0.098	0.098	0.098	0.098
	5%	0.049	0.049	0.049	0.049
	1%	0.010	0.010	0.010	0.010
0.2	10%	0.485	0.485	0.485	0.485
	5%	0.355	0.355	0.355	0.355
	1%	0.150	0.150	0.150	0.150
0.4	10%	0.919	0.919	0.919	0.919
	5%	0.871	0.871	0.871	0.871
	1%	0.655	0.655	0.655	0.655
0.6	10%	0.991	0.991	0.991	0.991
	5%	0.982	0.982	0.982	0.982
	1%	0.927	0.927	0.927	0.927
0.8	10%	0.999	0.999	0.999	0.999
	5%	0.998	0.998	0.998	0.998
	1%	0.987	0.987	0.987	0.987
1.0	10%	1.000	1.000	1.000	1.000
	5%	1.000	1.000	1.000	1.000
	1%	0.996	0.996	0.996	0.996

(1) R-B – ratio-based selection; $p(c_n) = \underset{1\leq l\leq p_v}{\arg\min} \dfrac{\widetilde{\lambda}_{l+1}}{\widetilde{\lambda}_l}$; references: Lam et al. (2011), Lam and Yao (2012), and Lee and Shao (2018, eqn. 6).

(2) Eι – entropy-based selection; $p(c_n) = \min\Big\{1 \leq l \leq p_v : \ E_l \geq \iota E_{p_v}\Big\}$; references: Alter et al. (2000) and Falini (2022, Sect. 2.6).

(3) TVι – Total Variance based selection; $p(c_n) = \displaystyle\sum_{l=1}^{p_v} \mathbb{1}\{\widetilde{f}_l \geq \iota\}$; references: Suhr (2005) and Falini (2022, Sect. 2.6).

(4) CTVι – Cumulative percentage of Total Variance based selection; $p(c_n) = \min\Big\{1 \leq l \leq p_v : \displaystyle\sum_{l'=1}^{l} \widetilde{f}_{l'} \geq \iota\Big\}$; references: Jolliffe (2002, Chapter 6) and Falini (2022, Sect. 2.6).

Recall $\widetilde{\Omega}_{h,n}$ is positive semi-definite hence the $\widetilde{f}_l$, $l = 1, \ldots, p_v$ are in descending order given a descending ordering of the eigenvalues $\widetilde{\lambda}_l$, $l = 1, \ldots, p_v$. This ensures that $p(c_n)$ per any of the above selection criteria corresponds to the largest $p(c_n)$ eigenvalues and the corresponding c_n is implicitly defined.

Table S.5 compares the size control and power performance of the χ^2-test with different choices of the regularization parameter c_n. Besides the ratio-based estimator, which fails to deliver a χ^2-test that controls size, the other choices lead to meaningful size control. One observes non-trivial power under $\mathbb{H}_a$. This exercise and that of Scenario 1 confirm the reliability and robustness of the selection rule $c_n = \widetilde{\lambda}_1 n^{-1/3}$ used in this paper.

S.5.5 Test of Nullity

$\mathbb{H}_o^*: \ \mathbb{E}[U \mid Z] = \mathbb{E}[U] = 0 \ a.s.$ is violated if either $\mathbb{E}[U] \neq 0$ or $\mathbb{P}\big(\mathbb{E}[U \mid Z] = \mathbb{E}[U]\big) < 1$. To compare the performance of the χ^2-test of the hypothesis of nullity $\mathbb{H}_o^*$, we take the following modified versions of DGP MI 6.

MI 6$'$: $U = 2\gamma\sqrt{3/5} + \mathcal{E}$, $Z \sim \mathcal{N}(0, 1)$ and $\mathcal{E} \sim \mathcal{U}[-\sqrt{3}, \sqrt{3}]$.

MI 6$''$: $U = 2\gamma\exp(-Z^2/3) + \mathcal{E}$, $Z \sim \mathcal{N}(0, 1)$ and $\mathcal{E} \sim \mathcal{U}[-\sqrt{3}, \sqrt{3}]$.

$\gamma \neq 0$ corresponds to $\mathbb{E}[U] \neq 0$ in MI 6$'$ and to $\mathbb{P}\big\{\mathbb{E}[U \mid Z] = \mathbb{E}[U]\big\} < 1$ in DGP MI 6$''$.
Table S.6 presents results on DGPs MI 6$'$ and MI 6$''$ using the framework of Table S.4 but with a focus on the power performance under either violation of $\mathbb{H}_o^*$. From both sets of columns corresponding to DGPs MI 6$'$ and MI 6$''$, one observes from Table S.6 that the χ^2-test of nullity has good size control and non-trivial power under violations of $\mathbb{E}[U] = 0$ or $\mathbb{E}[U \mid Z] = \mathbb{E}[U]$ $a.s.$

Table S.5: DGP MI 6 - SVD selection criteria

		$(c_n = \widetilde{\lambda}_1 n^{-\iota})$				Selection Criteria						
γ	Sig-Lev	$\iota = \frac{2}{5}$	$\iota = \frac{1}{3}$	$\iota = \frac{1}{4}$	$\iota = \frac{1}{6}$	R-B	E.7	E.9	TV.05	TV.10	CTV.7	CTV.9
0.0	10%	0.100	0.100	0.100	0.118	1.000	0.100	0.100	0.100	0.111	0.122	0.111
	5%	0.050	0.050	0.050	0.061	1.000	0.050	0.050	0.050	0.055	0.062	0.055
	1%	0.007	0.007	0.007	0.009	1.000	0.007	0.007	0.007	0.008	0.007	0.008
0.2	10%	0.973	0.973	0.973	0.981	1.000	0.973	0.973	0.973	0.979	0.988	0.979
	5%	0.945	0.945	0.945	0.963	1.000	0.945	0.945	0.945	0.958	0.971	0.958
	1%	0.822	0.822	0.822	0.853	1.000	0.822	0.822	0.823	0.841	0.901	0.841
0.4	10%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.6	10%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.8	10%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1.0	10%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	5%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Notes: The first four columns use the regularization technique used in this paper with $c_n = \widetilde{\lambda}_1 n^{-\iota}$, $\iota \in \{2/5, 1/3, 1/4, 1/6\}$, respectively. R-B is the ratio-based selection criterion, Eι, $\iota \in \{.7, .9\}$ is the α fraction of total entropy selection criterion, TVι, $\iota \in \{.05, .10\}$ is the α of total variance selection criterion, CTVι, $\iota \in \{.7, .9\}$ is the cumulative percentage of the total variance selection criterion.

Table S.6: DGPs MI 6$'$ and MI 6$''$

		DGP MI 6$'$, $c_n = \widetilde{\lambda}_1 n^{-\iota}$			
γ	Sig-Lev	$\iota = \frac{2}{5}$	$\iota = \frac{1}{3}$	$\iota = \frac{1}{4}$	$\iota = \frac{1}{6}$
0.0	10%	0.094	0.094	0.094	0.094
	5%	0.049	0.049	0.049	0.049
	1%	0.008	0.008	0.008	0.008
0.2	10%	1.000	1.000	1.000	1.000
	5%	1.000	1.000	1.000	1.000
	1%	0.998	0.998	0.998	0.998
0.4	10%	1.000	1.000	1.000	1.000
	5%	1.000	1.000	1.000	1.000
	1%	1.000	1.000	1.000	1.000
0.6	10%	1.000	1.000	1.000	1.000
	5%	1.000	1.000	1.000	1.000
	1%	1.000	1.000	1.000	1.000
0.8	10%	1.000	1.000	1.000	1.000
	5%	1.000	1.000	1.000	1.000
	1%	1.000	1.000	1.000	1.000
1.0	10%	1.000	1.000	1.000	1.000
	5%	1.000	1.000	1.000	1.000
	1%	1.000	1.000	1.000	1.000

		DGP MI 6$''$, $c_n = \widetilde{\lambda}_1 n^{-\iota}$			
γ	Sig-Lev	$\iota = \frac{2}{5}$	$\iota = \frac{1}{3}$	$\iota = \frac{1}{4}$	$\iota = \frac{1}{6}$
0.0	10%	0.094	0.094	0.094	0.094
	5%	0.049	0.049	0.049	0.049
	1%	0.008	0.008	0.008	0.008
0.2	10%	1.000	1.000	1.000	1.000
	5%	1.000	1.000	1.000	1.000
	1%	0.998	0.998	0.998	0.998
0.4	10%	1.000	1.000	1.000	1.000
	5%	1.000	1.000	1.000	1.000
	1%	1.000	1.000	1.000	1.000
0.6	10%	1.000	1.000	1.000	1.000
	5%	1.000	1.000	1.000	1.000
	1%	1.000	1.000	1.000	1.000
0.8	10%	1.000	1.000	1.000	1.000
	5%	1.000	1.000	1.000	1.000
	1%	1.000	1.000	1.000	1.000
1.0	10%	1.000	1.000	1.000	1.000
	5%	1.000	1.000	1.000	1.000
	1%	1.000	1.000	1.000	1.000

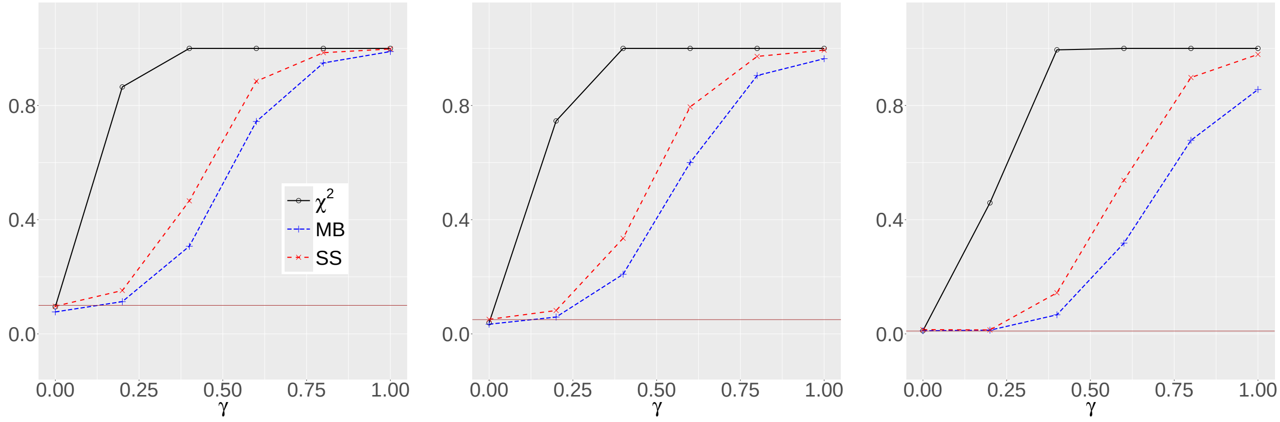
S.6 Monte Carlo Experiments - Specification Test - Non-linear Models

In this section we extend the simulation results in the main text to non-linear models. Specifically, we consider the problem of specification testing of propensity scores as studied in Sant'Anna and Song (2019). The treatment variable is generated as

$$\text{DGP NLM: } Y = \mathbb{1}\Big\{\theta_c + \sum_{l=1}^{5}\big(X_l\theta_l + \gamma(X_l^2 - 1)\big) \geq U\Big\},$$

where $X = Z$. Z and the parameters $[\theta_c, \theta_1, \ldots, \theta_5]^\top$ are specified as in Section 4 of the main text. U follows the standard logistic distribution. The logit model is estimated via maximum likelihood. To mitigate the impact of scale differences—often substantial in non-linear models—U and $h(Z)$ are normalized to have standard deviations of one and two, respectively, in constructing V_h.[2] This normalization ensures the procedure is scale-invariant. In the following simulation results, the proposed χ^2-test is compared to the Multiplier Bootstrap test of Escanciano (2024) and the propensity score specification test of Sant'Anna and Song (2019) (SS).

Figure S.9: DGP NLM – Gaussian Kernel – $n = 400$.



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

From the results above in Figure S.9, all tests demonstrate reasonable size control and non-trivial power under the alternative. Quite importantly, the proposed χ^2-test continues to perform in this non-linear setting with a binary limited dependent outcome. Observe that the wild bootstrap-based ICM

---

[2] This roughly preserves the scale ratio used in the linear models.

specification test of, e.g., Su and Zheng (2017) is not directly applicable in this case since the wild bootstrap therein cannot replicate outcomes that retain the two-point support of the binary outcome.

S.7 Relation to CM tests

The proposed test is rooted in ICM tests, but it also shares the advantages of CM tests (Newey, 1985; Tauchen, 1985), which are powerful if prior information on $\mathbb{H}_a$ is available. For example, if $\mathbb{E}[U \mid Z]$ can only take certain types of alternatives $f_1(Z), \cdots, f_{p_f}(Z)$, $p_f \geq 1$, then setting weight functions in CM tests along the span of these alternatives may yield optimal power (Newey, 1985). Such power enhancement is also allowed in the proposed test by augmenting V with a vector-valued function of Z. In the case of the bivariate $V_h = [h(Z), U - h(Z)]^\top$ in Lemma 2.1, power enhancement is also achievable by using $h(Z)$ to target alternatives. CM tests, which are closely related to the proposed test, are based on estimates of the form

$$\mathcal{T}_n^{CM} = \sum_{i=1}^{n} \widetilde{m}(Z_i)\widetilde{U}_i,$$

where $\widetilde{m}(\cdot)$ is a vector of non-degenerate weight functions. Although one may argue that $m_{\widetilde{V}}(Z) := \mathbb{E}\big[K(Z, Z^\dagger)(V_h^\dagger - \mathbb{E}V_h) \mid Z\big]$ in our case plays a role similar to $\widetilde{m}(Z_i)$ in CM tests, there are fundamental differences.

First, CM tests are not omnibus for any $\widetilde{m}(Z)$ of fixed dimension. There always exist certain forms $f_1(Z), \cdots, f_{p_g}(Z)$ in $\mathbb{E}[U \mid Z]$ under $\mathbb{H}_a$ such that $\mathcal{T}_n^{CM}$ has no power; this occurs when U is orthogonal to $\widetilde{m}(Z)$ under $\mathbb{H}_a$. This drawback drew much criticism from the literature and may have triggered the rapid development of ICM tests, see, e.g. Bierens (1982, 1990), Bierens and Ploberger (1997), and Delgado et al. (2006). Although the omnibus property for CM tests can be approximately attained by increasing the dimension of $\widetilde{m}(Z_i)$ via non-parametric techniques such as kernel smoothing (which is effectively what non-parametric tests do, e.g., Wooldridge (1992), Yatchew (1992), and Zheng (1996)), our proposed specification test remains omnibus with the dimension of V fixed; the proposed χ^2-test can therefore be viewed as a *consistent* CM test.

Second, our specification test allows V to be linearly dependent on U but CM tests do not. This also distinguishes our test from CM tests as ICM$(U \mid Z)$ is key to justifying the omnibus property of our test, see the proof of Lemma 2.1. Two independent copies (U, V, Z) and $(U^\dagger, V^\dagger, Z^\dagger)$ are jointly included in δ_h while $\mathbb{E}[\mathcal{T}_n^{CM}] = \mathbb{E}[\widetilde{m}(Z)U]$ involves only a single copy. If U is linearly included in the construction of $\widetilde{m}(\cdot)$, then most likely $\mathcal{T}_n^{CM}$ is non-null even under $\mathbb{H}_o$. A common feature shared by the proposed test and CM tests is the pivotal limiting distribution of the test statistic. This is achieved thanks to the non-degeneracy of $\widetilde{m}(Z)$ and $m_{\widetilde{V}}(Z)$.

References

Alter, O., Brown, P. O., & Botstein, D. (2000). Singular value decomposition for genome-wide expression data processing and modeling. *Proceedings of the National Academy of Sciences*, *97*(18), 10101–10106.

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*, 1443–1458.

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

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

Dufour, J.-M., & Valéry, P. (2016). *Rank-robust wald-type tests: A regularization approach* (tech. rep.). Working paper, McGill University/HEC.

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

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

Falini, A. (2022). A review on the selection criteria for the truncated SVD in data science applications. *Journal of Computational Mathematics and Data Science*, *5*, 100064.

Jolliffe, I. T. (2002). *Principal component analysis for special types of data*. Springer.

Lam, C., & Yao, Q. (2012). Factor modeling for high-dimensional time series: Inference for the number of factors. *The Annals of Statistics*, 694–726.

Lam, C., Yao, Q., & Bathia, N. (2011). Estimation of latent factors for high-dimensional time series. *Biometrika*, *98*(4), 901–918.

Lee, A. J. (1990). *U-statistics: Theory and practice*. Routledge.

Lee, C. E., & Shao, X. (2018). Martingale difference divergence matrix and its application to dimension reduction for stationary multivariate time series. *Journal of the American Statistical Association*, *113*(521), 216–229.

Maesono, Y. (1998). Asymptotic comparisons of several variance estimators and their effects for studentizations. *Annals of the Institute of Statistical Mathematics*, *50*(3), 451–470.

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

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

Seri, R. (2022). Computing the asymptotic distribution of second-order U-and V-statistics. *Computational Statistics & Data Analysis*, *174*, 107437.

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.

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

Suhr, D. D. (2005). Principal component analysis vs. exploratory factor analysis. *SUGI 30 proceedings*, *203*(230), 1–11.

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

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.

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

Zolotarev, V. M. (1961). Concerning a certain probability problem. *Theory of Probability & Its Applications*, *6*(2), 201–204.