

Feasible IV Regression without Excluded Instruments: Online supplement

EMMANUEL SELORM TSYAWO[†]

[†]*AIRESS & FGSES, Université Mohammed VI Polytechnique, Technopolis-Rabat. Morocco.*

E-mail: emmanuel.tsyawo@um6p.ma

Summary

The relevance condition of Integrated Conditional Moment (ICM) estimators is significantly weaker than the conventional IV's in at least two respects: (1) consistent estimation without excluded instruments is possible, provided endogenous covariates are non-linearly mean-dependent on exogenous covariates, and (2) endogenous covariates may be uncorrelated with but mean-dependent on instruments. These remarkable properties notwithstanding, multiplicative-kernel ICM estimators suffer diminished identification strength, large bias, and severe size distortions even for a moderately sized instrument vector. This paper proposes a computationally fast linear ICM estimator that better preserves identification strength in the presence of multiple instruments and a test of the ICM relevance condition. Monte Carlo simulations demonstrate a considerably better size control in the presence of multiple instruments and a favourably competitive performance in general. An empirical example illustrates the practical usefulness of the estimator, where estimates remain plausible when no excluded instrument is used.

Keywords: *endogeneity, martingale difference divergence, integrated conditional moment, linear completeness*

S1. PROOFS OF THEORETICAL RESULTS

S1.1. Proof of Theorem 3.2

By Assumption 3.3(b), Proposition 3.1(b), and (S.4), $-E[h(Z)'X] = -E[||Z - Z^\dagger||X'X^\dagger]$ is positive definite hence for some constant $\underline{\delta} > 0$ such that $2\underline{\delta} \leq \rho_{\min}(-E[||Z - Z^\dagger||X'X^\dagger])$ where $\rho_{\min}(\cdot)$ denotes the minimum eigen-value,

$$\begin{aligned} \underline{\delta} < \rho_{\min}(-E[||Z - Z^\dagger||X'X^\dagger]) &= \inf_{\tau \in \mathcal{S}_{p_x}} -\tau' E[||Z - Z^\dagger||X'X^\dagger] \tau \\ &= \inf_{\tau \in \mathcal{S}_{p_x}} -E[||Z - Z^\dagger|| (X\tau)(X^\dagger\tau)] \\ &\equiv \inf_{\tau \in \mathcal{S}_{p_x}} \overline{\text{MDD}}^2(X\tau|Z). \end{aligned}$$

Assumption 3.3(b) is thus equivalent to $\inf_{\tau \in \mathcal{S}_{p_x}} \overline{\text{MDD}}^2(X\tau|Z) > 0$ thanks to Proposition 3.1(b). Notice that $\tau^* \equiv \arg \inf_{\tau \in \mathcal{S}_{p_x}} \overline{\text{MDD}}^2(X\tau|Z)$ is the eigen-vector associated with the smallest eigen-value of $-E[||Z - Z^\dagger||X'X^\dagger]$. A test of Assumption 3.3(b) can thus be formulated as

$$\begin{aligned} \mathbb{H}'_o &: \overline{\text{MDD}}^2(X\tau^*|Z) = 0 \\ \mathbb{H}'_a &: \overline{\text{MDD}}^2(X\tau^*|Z) > 0. \end{aligned}$$

Assumption 3.3(b) fails under the null hypothesis \mathbb{H}'_o .

Partition τ^* conformably as $[\tau_1^*, \tau_{-1}^*]'$. From Lemma S.6, $\tau_1^* \neq 0$ under \mathbb{H}'_o . For $\eta^* \equiv -\tau_{-1}^*/\tau_1^*$,

$$\begin{aligned} \overline{\text{MDD}}^2(X\tau^*|Z) &= \overline{\text{MDD}}^2(\tau_1^*D + \tilde{X}\tau_{-1}^*|Z) = -E[\|Z - Z^\dagger\|(\tau_1^*D + \tilde{X}\tau_{-1}^*)(\tau_1^*D + \tilde{X}\tau_{-1}^*)^\dagger] \\ &= -(\tau_1^*)^2 E[\|Z - Z^\dagger\|(D + \tilde{X}\tau_{-1}^*/\tau_1^*)(D + \tilde{X}\tau_{-1}^*/\tau_1^*)^\dagger] \\ &= -(\tau_1^*)^2 E[\|Z - Z^\dagger\|(D - \tilde{X}\eta^*)(D - \tilde{X}\eta^*)^\dagger] = (\tau_1^*)^2 \overline{\text{MDD}}^2(D - \tilde{X}\eta^*|Z). \end{aligned}$$

By Lemma S.6, $\tau_1^* \neq 0$ under \mathbb{H}'_o thus $\overline{\text{MDD}}^2(X\tau^*|Z) = 0$ if and only if $\overline{\text{MDD}}^2(D - \tilde{X}\eta^*|Z) = 0$. Moreover, the parameter space of η^* is unconstrained hence the hypotheses can be reformulated as

$$\begin{aligned} \mathbb{H}_o &: \overline{\text{MDD}}^2(\mathcal{E}^D(\eta^*)|Z) = 0 \\ \mathbb{H}_a &: \overline{\text{MDD}}^2(\mathcal{E}^D(\eta^*)|Z) > 0. \end{aligned}$$

Under Assumption 3.2(a), the conclusion follows from Property (b) noting that $\overline{\text{MDD}}^2(\mathcal{E}^D(\eta^*)|Z) = 0$ is equivalent to $\text{MDD}^2(\mathcal{E}^D(\eta^*)|Z) = 0$ and $\overline{\text{MDD}}^2(\mathcal{E}^D(\eta^*)|Z) \geq 0$ by construction. \square

S1.2. Useful Lemmata

This section collects lemmata used in the proof of Theorems 3.1 and 3.2. Intermediate results for the proof of Theorem 3.1(a) are collected in lemmata S.1, S.2, and S.3. The following lemma derives the expression for B and establishes its boundedness.

LEMMA S.1. *Under Assumptions 3.1, 3.2, and 3.3(a), (a) $B = E[\sigma^2(Z)h(Z)'h(Z)]$; (b) $\|B\| \leq 4\bar{\sigma}^2C$.*

Proof:

Part (a):

$$B \equiv \text{var}[\sqrt{n}E_n[h(Z_i)'U_i]] = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n E[U_i U_j h(Z_i)'h(Z_j)] = E[U^2 h(Z)'h(Z)]$$

since for $i \neq j$, $E[U_i U_j h(Z_i)'h(Z_j)] = E[h(Z_i)'U_i]E[h(Z_j)'U_j]' = 0$ by independent and identical sampling (Assumption 3.1) and the exogeneity condition Assumption 3.3(a). For $i = j$, Assumption 3.1 and the Law of Iterated Expectations (LIE) imply $E[U_i^2 h(Z_i)'h(Z_i)] = E[E[U^2|Z]h(Z)'h(Z)] \equiv E[\sigma^2(Z)h(Z)'h(Z)]$ for each $i \in \{1, \dots, n\}$.

Part (b): For each $z \in \mathbb{R}^{p_z}$ defined on the support of Z ,

$$\begin{aligned} \|h(z)\| &= \|E[\|z - Z\|X]\| \\ &\leq \|z\|E\|X\| + E[\|Z\| \cdot \|X\|] \\ &\leq \|z\|(E\|X\|^4)^{1/4} + (E\|Z\|^4 \cdot E\|X\|^4)^{1/4} \\ &\leq \|z\|C^{1/4} + C^{1/2} \end{aligned} \tag{S.1}$$

by the expectation, triangle, Lyapunov, and Cauchy-Schwartz inequalities. The last inequality follows from Assumption 3.2(a).

Applying the c_r -inequality to (S.1) gives the bound

$$\|h(z)\|^2 \leq 2\|z\|^2 C^{1/2} + 2C. \quad (\text{S.2})$$

By the expectation inequality, Schwartz inequality, Assumption 3.2(b), (S.2), the Lyapunov inequality, and Assumption 3.2(a),

$$\begin{aligned} \|B\| &= \|E[\sigma^2(Z)h(Z)'h(Z)]\| \leq E[\sigma^2(Z)\|h(Z)\|^2] \leq \bar{\sigma}^2 E[\|h(Z)\|^2] \\ &\leq \bar{\sigma}^2(2E\|Z\|^2 C^{1/2} + 2C) \leq \bar{\sigma}^2(2(E\|Z\|^4)^{1/2} C^{1/2} + 2C) \\ &\leq 4\bar{\sigma}^2 C. \end{aligned} \quad (\text{S.3})$$

□

The next lemma concerns the strong consistency of the matrix $\hat{A}_n \equiv -E_n[h_n(Z_i)'X_i]$.

LEMMA S.2. *Under Assumption 3.1 and Assumption 3.2(a), $\hat{A}_n \xrightarrow{a.s.} A$.*

Proof: First, $E[E_n[h_n(Z_i)'X_i]] = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n E[\|Z_i - Z_j\|X_j'X_i] = E[\|Z - Z^\dagger\|X^\dagger'X] = E[\|Z - Z^\dagger\|X^\dagger'X]$ by Assumption 3.1. Observe that

$$\begin{aligned} -E[\|Z - Z^\dagger\|X^\dagger'X] &= -E[\|Z - Z^\dagger\|X^\dagger'E[X|Z, Z^\dagger, X^\dagger]] \\ &= -E[\|Z - Z^\dagger\|X^\dagger'E[X|Z]] \equiv -E[\|Z - Z^\dagger\|X^\dagger'm(Z)] \\ &= -E[E[\|Z - Z^\dagger\|X^\dagger|Z]m(Z)] \equiv -E[h(Z)'m(Z)] \\ &= -E[h(Z)'X], \end{aligned} \quad (\text{S.4})$$

hence $E[E_n[h_n(Z_i)'X_i]] = E[h(Z)'X]$. The first equality proceeds by the LIE, the second by independence as $E[X|Z, Z^\dagger, X^\dagger] = E[X|Z]$, the third by the LIE cum conditioning theorem, and the fourth by the LIE cum conditioning theorem as $E[h(Z)'X] = E[h(Z)'E[X|Z]] \equiv E[h(Z)'m(Z)]$.

By the Cauchy-Schwartz, Schwartz, and c_r inequalities,

$$\begin{aligned} E[\|Z - Z^\dagger\| \cdot \|X^\dagger'X\|] &\leq (E[\|Z - Z^\dagger\|^2])^{1/2} (E[\|X^\dagger'X\|^2])^{1/2} \\ &\leq (E[\|Z - Z^\dagger\|^2])^{1/2} (E[\|X\|^2\|X^\dagger\|^2])^{1/2} \\ &\leq (2E[\|Z\|^2 + \|Z^\dagger\|^2])^{1/2} E[\|X\|^2] \\ &= 2(E[\|Z\|^2])^{1/2} E[\|X\|^2]. \end{aligned}$$

Apply the Lyapunov inequality and Assumption 3.2(a) to obtain the bound $E[\|Z - Z^\dagger\| \cdot \|X^\dagger'X\|] \leq 2(E[\|Z\|^4])^{1/4} (E[\|X\|^4])^{1/2} \leq 2C^{3/4}$. This implies $E[\tilde{A}_{k,l}] \leq 2C^{3/4}$ for each $k, l = 1, \dots, p_x$ where $\tilde{A}_{k,l}$ is the (k, l) 'th element of the $p_x \times p_x$ random matrix $-\|Z - Z^\dagger\|X^\dagger'X$.

The conclusion follows from the strong law of large numbers (SLLN) for U -statistics – see Hoeffding (1961). □

LEMMA S.3. *Under Assumptions 3.1, 3.2, and 3.3(a),*

$$E_n[(h_n(Z_i) - h(Z_i))'U_i] = O_p(n^{-1}).$$

Proof: The proof proceeds by showing that $(E_n[\tau'(h_n(Z_i) - h(Z_i))'U_i])^2 = O_p(n^{-2})$ for all $\tau \in \mathcal{S}_{p_x}$ where $\mathcal{S}_p \equiv \{\tau \in \mathbb{R}^p : \|\tau\| = 1\}$.

$$(E_n[\tau'(h_n(Z_i) - h(Z_i))'U_i])^2 = \frac{1}{n^2(n-1)^2} \sum_{i=1}^n \sum_{j \neq i} \sum_{i'=1}^n \sum_{j' \neq i'} \psi_{ij} \psi_{i'j'}$$

where $\psi_{ij} = \|Z_i - Z_j\| \tau' X_j' U_i - E[\|Z_i - Z_j\| \tau' X_j | Z_i]' U_i = \|Z_i - Z_j\| \tau' X_j' U_i - \tau' h(Z_i)' U_i$. There are four cases to consider:

Case A: $i \neq i'$ and $j \neq j'$, $(n(n-1)^2(n-2))$ instances).

In this case, $\mathcal{A} \equiv E[\psi_{ij} \psi_{i'j'}] = E[\psi_{ij}] \times E[\psi_{i'j'}] = (E[\psi])^2$ by independent and identical sampling (Assumption 3.1) where $\psi \equiv \|Z - Z^\dagger\| \tau' X^\dagger U - \tau' h(Z)' U$.

$$\begin{aligned} E[\psi] &= E[\|Z - Z^\dagger\| \tau' X^\dagger E[U|Z, Z^\dagger, X^\dagger]] - \tau' E[h(Z)' E[U|Z]] \\ &= E[\|Z - Z^\dagger\| \tau' X^\dagger E[U|Z]] - \tau' E[h(Z)' E[U|Z]] \\ &= 0 \end{aligned}$$

thus $\mathcal{A} = 0$. The first equality follows from the LIE, the second by independence as $E[U|Z, Z^\dagger, X^\dagger] = E[U|Z]$, and the last equality follows from Assumption 3.3(a).

Case B: $i = i'$ and $j \neq j'$, $(n(n-1)(n-2))$ instances).

In this case,

$$\begin{aligned} \mathcal{B} &\equiv E[\psi_{ij} \psi_{i'j'}] = E[\psi_{ij} \psi_{ij'}] \\ &= E[(\|Z_i - Z_j\| \tau' X_j' U_i - \tau' h(Z_i)' U_i) \times (\|Z_i - Z_{j'}\| \tau' X_{j'}' U_i - \tau' h(Z_i)' U_i)] \\ &= E[(\|Z - Z^\dagger\| \tau' X^\dagger U - \tau' h(Z)' U) \times (\|Z - Z^{\dagger\dagger}\| \tau' X^{\dagger\dagger} U - \tau' h(Z)' U)] \\ &= E[\|Z - Z^\dagger\| \tau' X^\dagger \|Z - Z^{\dagger\dagger}\| X^{\dagger\dagger} \tau U^2] - E[\|Z - Z^\dagger\| \tau' X^\dagger h(Z) \tau U^2] \\ &\quad - E[\tau' h(Z)' \|Z - Z^{\dagger\dagger}\| X^{\dagger\dagger} \tau U^2] + E[\tau' h(Z)' h(Z) \tau U^2] \equiv \mathcal{B}_1 - \mathcal{B}_2 - \mathcal{B}_3 + \mathcal{B}_4 \end{aligned}$$

where $[X, Z]$, $[X^\dagger, Z^\dagger]$, and $[X^{\dagger\dagger}, Z^{\dagger\dagger}]$ are *iid* copies. The summands are studied in turn below.

$$\begin{aligned} \mathcal{B}_1 &= E[\|Z - Z^\dagger\| \tau' X^\dagger \|Z - Z^{\dagger\dagger}\| X^{\dagger\dagger} \tau E[U^2|Z, Z^\dagger, X^\dagger, Z^{\dagger\dagger}]] \\ &= E[\sigma^2(Z) \|Z - Z^\dagger\| \tau' X^\dagger \|Z - Z^{\dagger\dagger}\| X^{\dagger\dagger} \tau] \\ &= E[\sigma^2(Z) E[(\|Z - Z^\dagger\| \tau' X^\dagger \|Z - Z^{\dagger\dagger}\| X^{\dagger\dagger} \tau) | Z]] \\ &= E[\sigma^2(Z) E[\|Z - Z^\dagger\| \tau' X^\dagger | Z] \times E[\|Z - Z^{\dagger\dagger}\| X^{\dagger\dagger} \tau | Z]] \\ &= E[\sigma^2 \tau' h(Z)' h(Z) \tau] \end{aligned}$$

The first equality follows by the LIE, the second by independence as $E[U^2|Z, Z^\dagger, X^\dagger, Z^{\dagger\dagger}] = E[U^2|Z] \equiv \sigma^2(Z)$, the third by the LIE with the conditioning theorem (see, e.g., Hansen (2021, Theorem 2.3)), and the fourth by independence conditional on Z .

By the LIE, independence, and the LIE with the conditioning theorem,

$$\begin{aligned} \mathcal{B}_2 &= E[\|Z - Z^\dagger\| \tau' X^\dagger \tau' h(Z)' E[U^2|Z, Z^\dagger, X^\dagger]] \\ &= E[\sigma^2(Z) \|Z - Z^\dagger\| \tau' X^\dagger h(Z) \tau] \\ &= E[\sigma^2(Z) E[\|Z - Z^\dagger\| \tau' X^\dagger | Z] h(Z) \tau] \\ &= E[\sigma^2(Z) \tau' h(Z)' h(Z) \tau]. \end{aligned}$$

It is easily verified that $\mathcal{B}_3 = E[\sigma^2(Z) \tau' h(Z)' h(Z) \tau]$ using the same sequence of argu-

ments as the foregoing. By the LIE, $\mathcal{B}_4 = E[\sigma^2(Z)\tau'h(Z)'h(Z)\tau]$. Observe that $\mathcal{B}_1 = \mathcal{B}_2 = \mathcal{B}_3 = \mathcal{B}_4$, thus, $\mathcal{B} = E[\psi_{ij}\psi_{ij'}] = 0$.

Case C: $i \neq i'$ and $j = j'$, $(n(n-1)^2)$ instances).

In this case,

$$\begin{aligned} \mathcal{C} &\equiv E[\psi_{ij}\psi_{i'j'}] = E[\psi_{ij}\psi_{ij}] \\ &= E[(\|Z_i - Z_j\|\tau'X_j'U_i - \tau'h(Z_i)'U_i) \times (\|Z_{i'} - Z_j\|\tau'X_j'U_{i'} - \tau'h(Z_{i'})'U_{i'})] \\ &= E[(\|Z - Z^\dagger\|\tau'X^\dagger U - \tau'h(Z)'U) \times (\|Z^{\dagger\dagger} - Z^\dagger\|\tau'X^{\dagger\dagger}U^{\dagger\dagger} - \tau'h(Z^{\dagger\dagger})'U^{\dagger\dagger})] \\ &= E[\|Z - Z^\dagger\|\tau'X^\dagger\|Z^{\dagger\dagger} - Z^\dagger\|X^\dagger\tau U U^{\dagger\dagger}] - E[\|Z - Z^\dagger\|\tau'X^\dagger\tau h(Z^{\dagger\dagger})'U U^{\dagger\dagger}] \\ &\quad - E[\tau'h(Z)' \|Z^{\dagger\dagger} - Z^\dagger\| \tau'X^{\dagger\dagger}U U^{\dagger\dagger}] + E[\tau'h(Z)' h(Z^{\dagger\dagger})\tau U U^{\dagger\dagger}] \equiv \mathcal{C}_1 - \mathcal{C}_2 - \mathcal{C}_3 + \mathcal{C}_4. \end{aligned}$$

Define $\mu(z) = E[U|Z = z]$. By the LIE and independence,

$$\begin{aligned} \mathcal{C}_1 &= E[\|Z - Z^\dagger\|\tau'X^\dagger\|Z^{\dagger\dagger} - Z^\dagger\|X^\dagger\tau E[U U^{\dagger\dagger}|Z, Z^\dagger, X^\dagger, Z^{\dagger\dagger}]] \\ &= E[\|Z - Z^\dagger\| \cdot \|Z^{\dagger\dagger} - Z^\dagger\| \tau'X^\dagger X^\dagger \tau E[U U^{\dagger\dagger}|Z, Z^{\dagger\dagger}]] \\ &= E[\mu(Z)\mu(Z^{\dagger\dagger})\|Z - Z^\dagger\| \cdot \|Z^{\dagger\dagger} - Z^\dagger\| \tau'X^\dagger X^\dagger \tau]; \\ \mathcal{C}_2 &= E[\|Z - Z^\dagger\|\tau'X^\dagger h(Z^{\dagger\dagger})\tau E[U U^{\dagger\dagger}|Z, Z^\dagger, X^\dagger, Z^{\dagger\dagger}]] \\ &= E[\|Z - Z^\dagger\|\tau'X^\dagger h(Z^{\dagger\dagger})\tau E[U U^{\dagger\dagger}|Z, Z^{\dagger\dagger}]] \\ &= E[\mu(Z)\mu(Z^{\dagger\dagger})\|Z - Z^\dagger\|\tau'X^\dagger h(Z^{\dagger\dagger})\tau]; \\ \mathcal{C}_3 &= E[\tau'h(Z)' \|Z^{\dagger\dagger} - Z^\dagger\| X^\dagger \tau E[U U^{\dagger\dagger}|Z, Z^{\dagger\dagger}, Z^\dagger, X^\dagger]] \\ &= E[\tau'h(Z)' \|Z^{\dagger\dagger} - Z^\dagger\| X^\dagger \tau E[U U^{\dagger\dagger}|Z, Z^{\dagger\dagger}]] \\ &= E[\mu(Z)\mu(Z^{\dagger\dagger})\tau'h(Z)' \|Z^{\dagger\dagger} - Z^\dagger\| X^\dagger \tau]; \text{ and} \\ \mathcal{C}_4 &= E[\tau'h(Z)' h(Z^{\dagger\dagger})\tau E[U U^{\dagger\dagger}|Z, Z^{\dagger\dagger}]] \\ &= E[\mu(Z)\mu(Z^{\dagger\dagger})\tau'h(Z)' h(Z^{\dagger\dagger})\tau]. \end{aligned}$$

As $\mu(Z) = 0$ a.s. by Assumption 3.3(a), $\mathcal{C}_1 = \mathcal{C}_2 = \mathcal{C}_3 = \mathcal{C}_4 = 0$ thus $\mathcal{C} = 0$.

Case D: $i = i'$ and $j = j'$, $(n(n-1))$ instances).

In this case,

$$\begin{aligned} \mathcal{D} &\equiv E[\psi_{ij}\psi_{i'j'}] = E[\psi_{ij}^2] \\ &= E[(\|Z_i - Z_j\|\tau'X_j'U_i - \tau'h(Z_i)'U_i) \times (\|Z_i - Z_j\|\tau'X_j'U_i - \tau'h(Z_i)'U_i)] \\ &= E[(\|Z - Z^\dagger\|\tau'X^\dagger U - \tau'h(Z)'U) \times (\|Z - Z^\dagger\|\tau'X^\dagger U - \tau'h(Z)'U)] \\ &= E[\|Z - Z^\dagger\|^2 \tau'X^\dagger X^\dagger \tau U^2] - E[\|Z - Z^\dagger\|\tau'X^\dagger h(Z)\tau U^2] \\ &\quad - E[\tau'h(Z)' \|Z - Z^\dagger\| X^\dagger \tau U^2] + E[\tau'h(Z)' h(Z)\tau U^2] \equiv \mathcal{D}_1 - \mathcal{D}_2 - \mathcal{D}_3 + \mathcal{D}_4. \end{aligned}$$

Observe that by the LIE, independence, and the LIE with the conditioning theorem,

$$\begin{aligned} \mathcal{D}_2 &= E[\|Z - Z^\dagger\|\tau'X^\dagger h(Z)\tau E[U^2|Z, Z^\dagger, X^\dagger]] \\ &= E[\|Z - Z^\dagger\|\tau'X^\dagger h(Z)\tau E[U^2|Z]] \\ &= E[\sigma^2(Z)\|Z - Z^\dagger\|\tau'X^\dagger h(Z)\tau] \\ &= E[\sigma^2(Z)\tau'h(Z)'h(Z)\tau] = \mathcal{D}_3. \end{aligned}$$

By the LIE,

$$\mathcal{D}_4 = E[\sigma^2(Z)\tau'h(Z)'h(Z)\tau].$$

As $\mathcal{D}_2 = \mathcal{D}_3 = \mathcal{D}_4$, $\mathcal{D} \equiv E[\psi_{ij}^2] \geq 0$, and $\mathcal{D}_2 > 0$, $0 \leq \mathcal{D} = \mathcal{D}_1 - \mathcal{D}_2 \leq \mathcal{D}_1$. Further,

$$\begin{aligned}
\mathcal{D}_1 &= E[\|Z - Z^\dagger\|^2 (X^\dagger \tau)^2 E[U^2 | Z, Z^\dagger, X^\dagger]] \\
&= E[\sigma^2(Z) \|Z - Z^\dagger\|^2 (X^\dagger \tau)^2] \\
&\leq \bar{\sigma}^2 E[\|Z - Z^\dagger\|^2 (X^\dagger \tau)^2] \\
&\leq 2\bar{\sigma}^2 E[(\|Z\|^2 + \|Z^\dagger\|^2) \|X^\dagger\|^2] \\
&= 2\bar{\sigma}^2 (E[\|X^\dagger\|^2] \cdot E[\|Z\|^2] + E[\|X^\dagger\|^2 \cdot \|Z^\dagger\|^2]) \\
&\leq 2\bar{\sigma}^2 ((E[\|X\|^4] \cdot E[\|Z\|^4])^{1/2} + (E[\|X\|^4] \cdot E[\|Z\|^4])^{1/2}) \\
&\leq 4\bar{\sigma}^2 C < \infty.
\end{aligned}$$

The first equality follows by the LIE and the second by independence. The first and second inequalities, respectively, follow from Assumption 3.2(b) and the c_r -inequality. The third equality follows from the independence of X^\dagger and Z , the third inequality follows from the Lyapunov and Cauchy-Schwartz inequalities, and the fourth inequality follows from Assumption 3.2(a).

Combining cases A, B, C, and D,

$$E[(E_n[\tau'(h_n(Z_i) - h(Z_i))' U_i])^2] = \frac{n(n-1)}{n^2(n-1)^2} \mathcal{D} \leq \frac{4\bar{\sigma}^2 C}{n(n-1)}$$

for all $\tau \in \mathcal{S}_{p_x}$. By the Markov inequality, $E_n[\tau'(h_n(Z_i) - h(Z_i))' U_i] = O_p(n^{-1})$ as claimed. \square

The following strong law of large numbers (SLLN) result on $h_n(Z_i)$ is established for each $i \in \{1, \dots, n\}$. It is used in conjunction with Lemma S.5 in the proof of Theorem 3.1(c).

LEMMA S.4. *Under Assumptions 3.1 and 3.2(a), $h_n(Z_i) \xrightarrow{a.s.} h(Z_i)$ as $n \rightarrow \infty$ for each $i \in \{1, \dots, n\}$.*

Proof: First, $E[h_n(Z_i) - h(Z_i)] = \frac{1}{n-1} \sum_{j \neq i}^n E[\|Z_i - Z_j\| X_j - E[\|Z_i - Z_j\| X_j | Z_i]] = 0$ by the LIE for each i . Second, for each $i \in \{1, \dots, n\}$,

$$E\|h_n(Z_i)\| \leq \frac{1}{n-1} \sum_{j \neq i}^n E[\|Z_i - Z_j\| \cdot \|X_j\|] = E[\|Z - Z^\dagger\| \cdot \|X^\dagger\|]$$

by the triangle inequality and *iid* sampling (Assumption 3.1). By the triangle inequality, that $[X^\dagger, Z^\dagger]$ is an *iid* copy of $[X, Z]$, the Cauchy-Schwartz inequality, and the Lyapunov inequalities,

$$E[\|Z - Z^\dagger\| \cdot \|X^\dagger\|] \leq E\|X\| \cdot E\|Z\| + E[\|X\| \cdot \|Z\|] \leq 2(E\|X\|^4)^{1/4} (E\|Z\|^4)^{1/4}.$$

Applying Assumption 3.2(a) and the triangle inequality yields the bound $E\|h_n(Z_i)\| \leq 2C^{1/2}$ for each $i \in \{1, \dots, n\}$. Conditional on Z_i , $i \neq j$, $\|Z_i - Z_j\| X_j$ are *iid* under Assumption 3.1. The conclusion then follows from the SLLN. \square

LEMMA S.5. *Under Assumptions 3.1-3.3, (a) $\max_{1 \leq j \leq n} |\hat{U}_j - U_j|^2 \cdot E_n[\|h(Z_i)\|^2] = o_p(1)$;*

(b) $\max_{1 \leq j \leq n} |\hat{U}_j - U_j| \cdot E_n[|U_i| \cdot \|h(Z_i)\|^2] = o_p(1)$; (c) $E_n[\hat{U}_i^2 \|h_n(Z_j) - h(Z_j)\| \cdot (\|h_n(Z_i)\| + \|h(Z_i)\|)] = o_p(1)$.

Proof: The integrability of X_i holds uniformly in i by Assumption 3.2(a), i.e., $E\|X\|^4 \leq C \implies X_i = o_p(n^{1/4})$ uniformly in i – see, e.g., Hansen (2021, Theorem 6.16). In addition to the Schwartz inequality and the \sqrt{n} -consistency of $\hat{\theta}_n$ (Theorem 3.1(a)), the following holds uniformly in i :

$$|\hat{U}_i - U_i| \leq \|X_i\| \cdot \|\hat{\theta}_n - \theta_o\| = o_p(n^{-1/4}). \quad (\text{S.5})$$

Part (a): (S.3) implies $E[\|h(Z)\|^2] \leq 4C$ thus $E_n[\|h(Z_i)\|^2] = O_p(1)$. $\max_{1 \leq j \leq n} |\hat{U}_j - U_j|^2 = o_p(n^{-1/2})$ by (S.5) hence $\max_{1 \leq j \leq n} |\hat{U}_j - U_j|^2 \cdot E_n[\|h(Z_i)\|^2] = o_p(1)$ by Slutsky's theorem.

Part (b): By the LIE, the Lyapunov inequality, and (S.3), $E[|U| \cdot \|h(Z)\|^2] \leq E[(E[U^2|Z])^{1/2} \cdot \|h(Z)\|^2] \leq \bar{\sigma} E[\|h(Z)\|^2] \leq 4\bar{\sigma}C$. $E_n[|U_i| \cdot \|h(Z_i)\|^2]$ is thus $O_p(1)$ by the Markov inequality. Since $\max_{1 \leq j \leq n} |\hat{U}_j - U_j| = o_p(n^{-1/4})$ by (S.5), $\max_{1 \leq j \leq n} |\hat{U}_j - U_j| \cdot E_n[|U_i| \cdot \|h(Z_i)\|^2] = o_p(1)$ by Slutsky's theorem.

Part (c):

$\|h_n(Z_i)\| - \|h(Z_i)\| = o_p(1)$ for each i thanks to Lemma S.4 and the continuous mapping theorem. Thus, from parts (a) and (b) above,

$$\begin{aligned} \hat{U}_i^2 (\|h_n(Z_i)\| + \|h(Z_i)\|) &= (U_i^2 + (\hat{U}_i - U_i)^2 + 2U_i(\hat{U}_i - U_i)) \times (2\|h(Z_i)\| + o_p(1)) \\ &\leq 2U_i^2 \|h(Z_i)\| + U_i^2 \times o_p(1) + \max_{1 \leq j \leq n} |\hat{U}_j - U_j|^2 \cdot (2\|h(Z_i)\| + o_p(1)) \\ &\quad + 2 \max_{1 \leq j \leq n} |\hat{U}_j - U_j| \cdot |U_i| (2\|h(Z_i)\| + o_p(1)) = 2U_i^2 \|h(Z_i)\| + o_p(1). \end{aligned}$$

It follows from the LIE and (S.3) that $E[U^2 \|h(Z)\|^2] \leq 4\bar{\sigma}^2 C$, thus $\hat{U}_i^2 (\|h_n(Z_i)\| + \|h(Z_i)\|) = O_p(1)$. As $\|h_n(Z_i) - h(Z_i)\| = o_p(1)$ by Lemma S.4 for each $i \in \{1, \dots, n\}$, $E_n[\hat{U}_i^2 \|h_n(Z_j) - h(Z_j)\| \cdot (\|h_n(Z_i)\| + \|h(Z_i)\|)] = o_p(1)$ by Slutsky's theorem. \square

The following intermediate result is used in the proof of Theorem 3.2; it is useful in transforming the test into a specification test.

LEMMA S.6. *Under \mathbb{H}'_o , $\tau_1^* \neq 0$; the converse does not hold.*

Proof: The proof of the first part proceeds by contradiction. Suppose $\tau_1^* = 0$, then $\|\tau_{-1}^*\| = 1$ since $\tau^* \in \mathcal{S}_{p_x}$. This implies $\overline{\text{MDD}}^2(X\tau^*|Z) = \overline{\text{MDD}}^2(\tau_1^*D + \tilde{X}\tau_{-1}^*|Z) = \overline{\text{MDD}}^2(\tilde{X}\tau_{-1}^*|Z) > 0$ which violates \mathbb{H}'_o since Z contains exogenous covariates \tilde{X} thus $\tilde{X}\tau_{-1}^*$ is not mean independent of Z . Hence, $\tau_1^* \neq 0$ if \mathbb{H}'_o is true.

For the second part, it is argued that if $\tau_1^* \neq 0$, either \mathbb{H}'_o or \mathbb{H}'_a can hold. Two cases are distinguished below.

First, if $|\tau_1^*| = 1$, then $\tau_{-1}^* = 0$ and $\overline{\text{MDD}}^2(X\tau^*|Z) = \overline{\text{MDD}}^2(\tau_1^*D + \tilde{X}\tau_{-1}^*|Z) = \overline{\text{MDD}}^2(D|Z)$ which may be zero or positive depending on whether D is mean-independent of Z (if D is zero-mean) or not.

Second, if $|\tau_1^*| \in (0, 1)$, then $\|\tau_{-1}^*\| \neq 0$ and $\overline{\text{MDD}}^2(X\tau^*|Z) = \overline{\text{MDD}}^2(\tau_1^*D + \tilde{X}\tau_{-1}^*|Z) > 0$

since Z contains exogenous covariates \tilde{X} . \square

S1.3. Proofs of Propositions

Proof of Proposition 2.1: Expanding terms, the objective function (2.2) can be expressed as

$$\begin{aligned} Q_o(\theta) &= -E[\|Z - Z^\dagger\|Y^\dagger Y] + E[\|Z - Z^\dagger\|X^\dagger Y]\theta + E[\|Z - Z^\dagger\|XY^\dagger]\theta - \theta' E[\|Z - Z^\dagger\|X^\dagger X]\theta \\ &= -E[\|Z - Z^\dagger\|Y^\dagger Y] + 2E[\|Z - Z^\dagger\|X^\dagger Y]\theta - \theta' E[\|Z - Z^\dagger\|X^\dagger X]\theta. \end{aligned}$$

noting that $E[\|Z - Z^\dagger\|X^\dagger Y] = E[\|Z - Z^\dagger\|XY^\dagger]$. Taking the first derivative and equating it to zero gives

$$\begin{aligned} 2E[\|Z - Z^\dagger\|X^\dagger Y] - E[\|Z - Z^\dagger\|X^\dagger X]\theta - E[\|Z - Z^\dagger\|X^\dagger X]\theta \\ = 2E[\|Z - Z^\dagger\|X^\dagger Y] - 2E[\|Z - Z^\dagger\|X^\dagger X]\theta = 0 \end{aligned}$$

by the equality of $E[\|Z - Z^\dagger\|X^\dagger X]$ and $E[\|Z - Z^\dagger\|X^\dagger X]$ due to symmetry. Thus, the minimiser of (2.2) is given by

$$\theta_o = (E[\|Z - Z^\dagger\|X^\dagger X])^{-1} E[\|Z - Z^\dagger\|X^\dagger Y].$$

Applying the analogue principle, the analytical expression of the estimator is given by

$$\begin{aligned} \hat{\theta}_n &= (E_n[\|Z_i - Z_j\|X_j' X_i])^{-1} E_n[\|Z_i - Z_j\|X_j' Y_i] \\ &= \left[\frac{1}{n} \sum_{i=1}^n \left(\frac{1}{n} \sum_{j=1}^n \|Z_i - Z_j\|X_j \right)' X_i \right]^{-1} \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{n} \sum_{j=1}^n \|Z_i - Z_j\|X_j \right)' Y_i \\ &= \left[\frac{1}{n} \sum_{i=1}^n \left(\frac{1}{n-1} \sum_{j=1}^n \|Z_i - Z_j\|X_j \right)' X_i \right]^{-1} \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{n-1} \sum_{j=1}^n \|Z_i - Z_j\|X_j \right)' Y_i \\ &= (E_n[h_n(Z_i)' X_i])^{-1} E_n[h_n(Z_i)' Y_i]. \end{aligned}$$

\square

Proof of Proposition 3.1:

Part (a): By the data generating process (DGP) $Y = X\theta_o + U$ (Assumption 3.1), the expected value of the objective function (2.2) can be expressed as

$$\begin{aligned} Q_o(\theta) &= -E[\|Z - Z^\dagger\|(U - X(\theta - \theta_o))(U^\dagger - X^\dagger(\theta - \theta_o))] \\ &= -E[\|Z - Z^\dagger\|UU^\dagger] + 2E[\|Z - Z^\dagger\|X^\dagger U](\theta - \theta_o) - (\theta - \theta_o)' E[\|Z - Z^\dagger\|X^\dagger X](\theta - \theta_o) \\ &= -E[\|Z - Z^\dagger\|UU^\dagger] + 2E[h(Z)' U](\theta - \theta_o) - (\theta - \theta_o)' E[h(Z)' X](\theta - \theta_o) \\ &= -(\theta - \theta_o)' E[h(Z)' X](\theta - \theta_o). \end{aligned}$$

$E[h(Z)' X] = E[\|Z - Z^\dagger\|X^\dagger X]$ by (S.4), and $E[h(Z)' U] = E[\|Z - Z^\dagger\|X^\dagger U]$ follows from similar arguments. The last equality follows from Assumption 3.3(a) as $MDD^2(U|Z) \equiv -E[\|Z - Z^\dagger\|UU^\dagger] = 0$ and $E[h(Z)' U] = 0$. Note that $Q_o(\theta_o) = -E[\|Z - Z^\dagger\|UU^\dagger] = 0$ thus from the foregoing,

$$Q_o(\theta) - Q_o(\theta_o) = -(\theta - \theta_o)' E[h(Z)' X](\theta - \theta_o).$$

$-E[h(Z)'X]$ is positive definite under Assumption 3.3(b) – see (3.1). It follows that $Q_o(\theta) > Q_o(\theta_o)$ for all $\theta \neq \theta_o$, and θ_o is identified as claimed.

Part (b): The proof of Proposition 3.1(b) follows that of Proposition 2.1 in Escanciano (2018) and is thus omitted.

Part (c): Let \mathcal{T} denote any $p_x \times r$ matrix whose columns comprise $[\tau_1, \dots, \tau_r]$, where $\tau_t \in \mathcal{S}_{p_x}$ for each $t \in \{1, \dots, r\}$ without loss of generality. As by Property (b), $E[X\tau_t|Z] = 0$ is equivalent to $\overline{\text{MDD}}^2(X\tau_t|Z) = -\tau_t'E[||Z - Z^\dagger||X'X^\dagger]\tau_t = 0$ for each $t \in \{1, \dots, r^*\}$, $\text{rank}(E[||Z - Z^\dagger||X'X^\dagger]) = p_x - r^*$ if and only if there exists r^* linearly independent combinations of X that are mean-independent of Z .

Consider a matrix $E[Z'X]$ where there are r^\dagger linearly independent combinations of X that are uncorrelated with Z . Since mean independence implies zero correlation, it follows that $r^* \leq r^\dagger$, and $p_x \geq p_x - r^* = \text{rank}(E[||Z - Z^\dagger||X'X^\dagger]) \geq \text{rank}(E[Z'X]) = p_x - r^\dagger$. This completes the proof. \square

S1.4. Useful Propositions

The following propositions provide results underlying the discussion in Section 2.4.

PROPOSITION S.1. *Let ζ_M, ζ_G , and ζ_E be positive random variables which are appropriate means of measurable functions of $[Z, Z^\dagger]$ as outlined in Table 1, then for all $p_z \geq 1$, the following representations hold: (a) $K(Z, Z^\dagger) = -\zeta_M\sqrt{p_z}$ for the MMD kernel; (b) $K(Z, Z^\dagger) = \zeta_G^{p_z}$, $\zeta_G \in [0, 1]$ for the Gaussian kernel; and (c) $K(Z, Z^\dagger) \leq \zeta_E/\sqrt{p_z}$, $\zeta_E \in [0, 2\pi(\frac{\sqrt{\pi e}}{3})^3]$ for the ESC6 kernel.*

Proof:

Part (a): $K(Z, Z^\dagger) = -||Z - Z^\dagger|| = -\sqrt{p_z} \times ||(Z - Z^\dagger)/\sqrt{p_z}||$ is the MMD kernel. Set $\zeta_M = ||(Z - Z^\dagger)/\sqrt{p_z}|| = \sqrt{\frac{1}{p_z} \sum_{k=1}^{p_z} (Z_k - Z_k^\dagger)^2}$ where ζ_M is the quadratic mean of $\{(Z_k - Z_k^\dagger), 1 \leq k \leq p_z\}$.

Part (b): $K(Z, Z^\dagger) = \exp(-0.5||Z - Z^\dagger||^2) = \exp(-0.5p_z||Z - Z^\dagger||^2/p_z) = (\exp(-0.5||Z - Z^\dagger||^2/p_z))^{p_z}$ is the Gaussian kernel. Set $\zeta_G = \exp(-0.5||Z - Z^\dagger||^2/p_z) = \left(\prod_{k=1}^{p_z} \exp(-0.5(Z_k - Z_k^\dagger)^2)\right)^{1/p_z}$ where ζ_G is the geometric mean of $\{\exp(-0.5(Z_k - Z_k^\dagger)^2), 1 \leq k \leq p_z\}$.

Part (c): $K_n(Z_i, Z_j) = \frac{\pi^{(p_z/2)-1}}{\Gamma((p_z/2)+1)} \sum_{l=1}^n A_{ijl}^{(0)}/n$ is the ESC6 kernel.

First, observe that $A_{ijl}^{(0)} \in [0, \pi] \cup \{2\pi\}$ for any $i, j, l \in \{1, \dots, n\}$. This implies $\frac{1}{n} \sum_{l=1}^n A_{ijl}^{(0)} \in [0, 2\pi]$.¹

Second, by Stirling's formula $\Gamma(p+1) \approx \sqrt{2p\pi}(\frac{p}{e})^p$, $\Gamma(p_z/2+1) \approx \pi^{1/2}p_z^{1/2}(\frac{p_z}{2e})^{p_z/2}$ whence

$$\begin{aligned} \frac{\pi^{(p_z/2)-1}}{\Gamma((p_z/2)+1)} &\approx \frac{\pi^{(p_z/2)-1}}{\pi^{1/2}p_z^{1/2}(\frac{p_z}{2e})^{p_z/2}} = \left(\frac{2\pi e}{p_z}\right)^{p_z/2} \frac{1}{\pi^{3/2}\sqrt{p_z}} \\ &\leq \frac{1}{\pi^{3/2}\sqrt{p_z}} \times \max_{p \in \mathbb{N}} \left(\frac{2\pi e}{p}\right)^{p/2} = \left(\frac{\sqrt{\pi e}}{3}\right)^3 \frac{1}{\sqrt{p_z}} \end{aligned}$$

¹Recall $A_{ijl}^{(0)} = 2\pi$ if $Z_l = Z_i = Z_j$ else $A_{ijl}^{(0)} \in [0, \pi]$.

since $\left(\frac{2\pi e}{p}\right)^{p/2}$ attains its global maximum at $p = 6$ over the set of natural numbers \mathbb{N} .

It can be verified that for all $p_z \geq 1$, $\frac{\pi^{(p_z/2)-1}}{\Gamma((p_z/2)+1)} \leq \left(\frac{\sqrt{\pi e}}{3}\right)^3 \frac{1}{\sqrt{p_z}}$ thus

$$K_n(Z_i, Z_j) = \frac{\pi^{(p_z/2)-1}}{\Gamma((p_z/2)+1)} \sum_{l=1}^n A_{ijl}^{(0)}/n \leq \frac{\left(\frac{\sqrt{\pi e}}{3}\right)^3}{\sqrt{p_z}} \sum_{l=1}^n A_{ijl}^{(0)}/n \equiv \zeta_{n,E}/\sqrt{p_z}$$

where $\zeta_{n,E} \equiv \left(\frac{\sqrt{\pi e}}{3}\right)^3 \sum_{l=1}^n A_{ijl}^{(0)}/n$ is the arithmetic mean of $\left\{\left(\frac{\sqrt{\pi e}}{3}\right)^3 A_{ijl}^{(0)}, \leq l \leq n\right\}$. The conclusion follows from taking the almost sure limit $\zeta_{n,E} \xrightarrow{a.s.} \zeta_E$. \square

Let $F_Z(\cdot)$ be the cumulative distribution of Z and $Z \vee Z^\dagger \equiv [Z_1 \vee Z_1^\dagger, \dots, Z_{p_z} \vee Z_{p_z}^\dagger]$ denote the vector of element-wise maxima of Z and Z^\dagger . The following result derives the almost sure limit of the DL kernel.

PROPOSITION S.2. *Under Assumption 3.1, the DL kernel satisfies $K_n(Z_i, Z_j) \xrightarrow{a.s.} 1 - F_Z(Z_i \vee Z_j)$.*

Proof:

$$\mathbf{I}(Z_i \leq Z_l)\mathbf{I}(Z_j \leq Z_l) = \mathbf{I}(Z_l \geq Z_i)\mathbf{I}(Z_l \geq Z_j) = \mathbf{I}(Z_l \geq (Z_i \vee Z_j))$$

for any $i, j, l = 1, \dots, n$. Thus,

$$\begin{aligned} K_n(Z_i, Z_j) &= \frac{1}{n} \sum_{l=1}^n \mathbf{I}(Z_i \leq Z_l)\mathbf{I}(Z_j \leq Z_l) = \frac{1}{n} \sum_{l=1}^n \mathbf{I}(Z_l \geq (Z_i \vee Z_j)) \\ &= \left(\frac{n-2}{n}\right) \frac{1}{n-2} \sum_{l \neq i, j}^n \mathbf{I}(Z_l \geq (Z_i \vee Z_j)) + \frac{\mathbf{I}(Z_i \geq (Z_i \vee Z_j)) + \mathbf{I}(Z_j \geq (Z_i \vee Z_j))}{n} \\ &= \frac{1}{n-2} \sum_{l \neq i, j}^n \mathbf{I}(Z_l \geq (Z_i \vee Z_j)) + \frac{\mathbf{I}(Z_i \geq (Z_i \vee Z_j)) + \mathbf{I}(Z_j \geq (Z_i \vee Z_j))}{n} \\ &\quad - \frac{2}{n(n-2)} \sum_{l \neq i, j}^n \mathbf{I}(Z_l \geq (Z_i \vee Z_j)) \\ &\equiv \frac{1}{n-2} \sum_{l \neq i, j}^n \mathbf{I}(Z_l \geq (Z_i \vee Z_j)) + \nu_{ijn} \end{aligned}$$

Conditional on $[Z_i, Z_j]$, $\{\mathbf{I}(Z_l \geq (Z_i \vee Z_j)), l \neq i, j\}$ are *iid* under Assumption 3.1, and since the indicator function is bounded by construction,

$$\frac{1}{n-2} \sum_{l \neq i, j}^n \mathbf{I}(Z_l \geq (Z_i \vee Z_j)) \xrightarrow{a.s.} 1 - F_Z(Z_i \vee Z_j)$$

by the strong law of large numbers. By the boundedness of the indicator function, $|\nu_{ijn}| \leq 2/n$ for any Z_i, Z_j whence $\nu_{ijn} \xrightarrow{a.s.} 0$. Combining both steps above and Slutsky's theorem completes the proof. \square

PROPOSITION S.3. Let $Z \sim \mathcal{N}(0, I_{p_z})$ and Assumption 3.1 hold, then (a) $\text{var}[K(Z, Z^\dagger)] = 5^{-p_z/2} - 3^{-p_z}$ for the Gaussian kernel $K(Z, Z^\dagger) = \exp(-0.5(Z - Z^\dagger)'(Z - Z^\dagger))$; (b) $\text{var}[K(Z, Z^\dagger)] = 2p_z - 4\left(\frac{\Gamma((p_z+1)/2)}{\Gamma(p_z/2)}\right)^2$ for the MMD's kernel $K(Z, Z^\dagger) = -\|Z - Z^\dagger\|$; (c) $\lim_{n \rightarrow \infty} \text{var}[K_n(Z_i, Z_j)] = \text{var}[K(Z, Z^\dagger)] = 2^{-p_z} - (2/3)^{2p_z}$ for the almost sure limit of the DL kernel $K(Z, Z^\dagger) = 1 - F_Z(Z \vee Z^\dagger)$.

Proof:

Part (a):

$$K(Z, Z^\dagger)^2 = \exp(-(Z - Z^\dagger)'(Z - Z^\dagger)) = \prod_{k=1}^{p_z} \exp(-(Z_k - Z_k^\dagger)^2).$$

Notice that $(Z_k - Z_k^\dagger) \sim \mathcal{N}(0, 2)$ for each $k \in \{1, \dots, p_z\}$, thus by the independence and identical distribution of each element in Z ,

$$E[K(Z, Z^\dagger)^2] = \prod_{k=1}^{p_z} E[\exp(-(Z_k - Z_k^\dagger)^2)] = (E[\exp(-\varsigma_G^2)])^{p_z} \quad (\text{S.6})$$

where $\varsigma_G \sim \mathcal{N}(0, 2)$.

$$\begin{aligned} E[\exp(-\varsigma_G^2)] &= \frac{1}{2\sqrt{\pi}} \int_{-\infty}^{\infty} \exp(-\varsigma^2) \exp(-\frac{1}{4}\varsigma^2) d\varsigma = \frac{1}{2\sqrt{\pi}} \int_{-\infty}^{\infty} \exp(-\frac{5}{4}\varsigma^2) d\varsigma \\ &= \frac{1}{\sqrt{5}} \end{aligned}$$

using integration by substitution and that the probability density function of a normally distributed variable integrates to 1. It follows from (S.6) that $E[K(Z, Z^\dagger)^2] = 5^{-p_z/2}$.

By similar calculations, $E[K(Z, Z^\dagger)] = 3^{-p_z/2}$. Thus $\text{var}[K(Z, Z^\dagger)] = E[K(Z, Z^\dagger)^2] - (E[K(Z, Z^\dagger)])^2 = 5^{-p_z/2} - 3^{-p_z}$.

Part (b): $E[\|Z - Z^\dagger\|^2] = \sum_{k=1}^{p_z} E[(Z_k - Z_k^\dagger)^2] = p_z E[\varsigma_M^2] = 2p_z$ where $\varsigma_M \sim \mathcal{N}(0, 2)$. $\|Z - Z^\dagger\| = \sqrt{\sum_{k=1}^{p_z} (Z_k - Z_k^\dagger)^2} = \sqrt{\sum_{k=1}^{p_z} \varsigma_M^2}$ is chi-distributed with p_z degrees of freedom, a scale parameter $\sqrt{2}$, and $E[\|Z - Z^\dagger\|] = 2\frac{\Gamma((p_z+1)/2)}{\Gamma(p_z/2)}$ – see, e.g., Johnson et al. (1995, eqns 18.65 and 18.66). It thus follows that

$$\text{var}[-\|Z - Z^\dagger\|] = E[\|Z - Z^\dagger\|^2] - (E[\|Z - Z^\dagger\|])^2 = 2p_z - 4\left(\frac{\Gamma((p_z+1)/2)}{\Gamma(p_z/2)}\right)^2.$$

Part (c): From Proposition S.2, $K_n(Z_i, Z_j) = K(Z_i, Z_j) + o_p(1)$, thus $\text{var}[K_n(Z_i, Z_j)] = \text{var}[K(Z_i, Z_j)] + o(1)$ by the CMT. It follows from the *iid* assumption (Assumption 3.1) that

$$\lim_{n \rightarrow \infty} \text{var}[K_n(Z_i, Z_j)] = \text{var}[K(Z, Z^\dagger)] = \text{var}[1 - F_Z(Z \vee Z^\dagger)] = \text{var}[F_Z(Z \vee Z^\dagger)].$$

Under independence and normal distribution of the elements of Z , $F_Z(Z \vee Z^\dagger) = \prod_{k=1}^{p_z} \Phi(Z_k \vee Z_k^\dagger)$. By the identical distribution of the elements of Z and Z^\dagger ,

$$\begin{aligned} \text{var}[K(Z, Z^\dagger)] &= \text{var}\left[\prod_{k=1}^{p_z} \Phi(Z_k \vee Z_k^\dagger)\right] = E\left[\prod_{k=1}^{p_z} \Phi(Z_k \vee Z_k^\dagger)^2\right] - \left(E\left[\prod_{k=1}^{p_z} \Phi(Z_k \vee Z_k^\dagger)\right]\right)^2 \\ &= (E[\Phi(\varsigma_D)^2])^{p_z} - (E[\Phi(\varsigma_D)])^{2p_z} \end{aligned}$$

where the random ς_D denotes the maximum of two independent standard normally distributed variables. The cumulative distribution function of ς_D is given by $F_{\varsigma_D}(x) \equiv \mathbb{P}(\varsigma_D \leq x) = \Phi(x)^2$ hence the probability density function is $\frac{dF_{\varsigma_D}(x)}{dx} = 2\Phi(x)\phi(x)$.

Let $G(x)$ be a continuously differentiable function with first derivative $g(x) = \frac{dG(x)}{dx}$. For any positive integer k , integration by parts gives

$$\int G(x)^k g(x) dx = G(x)^{k+1} - k \int G(x)^k g(x) dx + \dot{C},$$

thus

$$\int G(x)^k g(x) dx = \frac{G(x)^{k+1}}{k+1} + \frac{\dot{C}}{k+1} \quad (\text{S.7})$$

where \dot{C} is a constant of integration.

It follows from (S.7) that

$$E[\Phi(\varsigma_D)^2] = \int_{-\infty}^{\infty} \Phi(\varsigma)^2 dF_{\varsigma_D}(\varsigma) = 2 \int_{-\infty}^{\infty} \Phi(\varsigma)^3 \phi(\varsigma) d\varsigma = 2 \frac{\Phi(\varsigma)^4}{4} \Big|_{-\infty}^{\infty} = \frac{1}{2}, \text{ and}$$

$$E[\Phi(\varsigma_D)] = \int_{-\infty}^{\infty} \Phi(\varsigma) dF_{\varsigma_D}(\varsigma) = 2 \int_{-\infty}^{\infty} \Phi(\varsigma)^2 \phi(\varsigma) d\varsigma = 2 \frac{\Phi(\varsigma)^3}{3} \Big|_{-\infty}^{\infty} = \frac{2}{3}.$$

Substituting the above values into the expression of $\text{var}[K(Z, Z^\dagger)]$ completes the proof. \square

S2. THE GENERALISED MARTINGALE DIFFERENCE CORRELATION (GMDC)

The definition of the GMDC follows Shao and Zhang (2014, p. 1304) which uses the \mathcal{D} -centring approach of Székely et al. (2007). Given an *iid* sample $\{[W_i, Z_i], 1 \leq i \leq n\}$, define the following: (1) $a_{kl} = W_k W_l$, $\bar{a}_{k\cdot} = W_k E_n[W_i]$, $\bar{a}_{\cdot l} = E_n[W_i] W_l$, and $\bar{a}_{\cdot\cdot} = (E_n[W_i])^2$; (2) $b_{kl} = K(Z_k, Z_l)$, $\bar{b}_{k\cdot} = n^{-1} \sum_{l=1}^n b_{kl}$, $\bar{b}_{\cdot l} = n^{-1} \sum_{k=1}^n b_{kl}$, and $\bar{b}_{\cdot\cdot} = n^{-2} \sum_{k=1}^n \sum_{l=1}^n b_{kl}$; and (3) $A_{kl} = a_{kl} - \bar{a}_{k\cdot} - \bar{a}_{\cdot l} + \bar{a}_{\cdot\cdot}$, and $B_{kl} = b_{kl} - \bar{b}_{k\cdot} - \bar{b}_{\cdot l} + \bar{b}_{\cdot\cdot}$. Then, the GMDC is computed as

$$\text{GMDC}_n(W|Z) = \frac{n^{-2} \sum_{i=1}^n \sum_{j=1}^n A_{ij} B_{ij}}{\sqrt{\left(n^{-2} \sum_{i=1}^n \sum_{j=1}^n A_{ij}^2 \right) \left(n^{-2} \sum_{i=1}^n \sum_{j=1}^n B_{ij}^2 \right)}}.$$

S3. MONTE CARLO SIMULATIONS - SUPPLEMENT

S3.1. Bias and size control, $n = 250$

For comparison with DGPs examined in the literature on ICM estimators, this subsection examines DGP_{2A} and DGP_{2B} taken from Antoine and Lavergne (2014) and DGP_{3A} and DGP_{3B} taken from Escanciano (2018). Define the function $f_2(Z) \equiv \sqrt{\frac{2\pi\sqrt{27}}{p_z} \sum_{k=1}^{p_z} Z_k \exp(-Z_k^2/2)}$. Heteroskedasticity is modelled in some specifications via the skedastic function $s(Z) = \sqrt{\left(1 + \frac{1}{p_z} \sum_{k=1}^{p_z} Z_k^2\right)/2}$.

$$DGP_{2A} : \begin{cases} Y = \alpha_o + \beta_o D + s(Z)U \\ D = \sqrt{\delta}Z/n^\delta + V \end{cases}, DGP_{2B} : \begin{cases} Y = \alpha_o + \beta_o D + s(Z)U \\ D = \sqrt{\delta}f_2(Z)/n^\delta + V \end{cases}$$

$$DGP_{3A} : \begin{cases} Y = \alpha_o + \beta_o D + U \\ D = \delta Z + V \end{cases}, DGP_{3B} : \begin{cases} Y = \alpha_o + \beta_o D + U \\ D = I(1 + \delta Z + V > 0) \end{cases}$$

DGP_{2A} and DGP_{2B} serve to examine performance under linear and non-linear “semi-strong” identification respectively – see Antoine and Lavergne (2014). DGP_{3A} and DGP_{3B} enable a comparison of estimators in the ICM- and K-classes under continuous and binary forms of non-parametric identification. For the parameters, $\alpha_o = \beta_o = 0$ in DGP_{2A} and DGP_{2B} , and $\alpha_o = \beta_o = \gamma_o = 1$ in DGP_{3A} and DGP_{3B} . As in Antoine and Lavergne (2014) and Escanciano (2018), $\rho = 0.8$ in DGP_{2A} , DGP_{2B} , DGP_{3A} , and DGP_{3B} . $p_z = 1$ in DGP_{3A} and DGP_{3B} . Following Antoine and Lavergne (2014), $p_z = 1$ for the ICM-class of estimators and the authors’ approach is used to generate 12 instruments for the K-class in DGP_{2A} and DGP_{2B} .²

Table S.1: DGP_{2A} , $n = 250$

	$\delta = 0.45$				$\delta = 0.30$				$\delta = 0.15$			
	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.
MMD	-0.015	0.202	0.391	0.081	-0.003	0.092	0.145	0.051	-0.001	0.042	0.062	0.045
WMD	-0.011	0.213	2.574	0.058	-0.016	0.092	0.147	0.047	-0.003	0.041	0.061	0.043
WMDF	-0.211	0.211	3.138	0.060	-0.015	0.092	0.146	0.047	-0.002	0.041	0.061	0.043
DL	-0.029	0.200	0.991	0.070	-0.005	0.089	0.141	0.053	-0.001	0.040	0.061	0.048
ESC6	-0.025	0.192	0.349	0.071	-0.005	0.088	0.141	0.052	-0.001	0.039	0.061	0.048
IIV	-0.003	0.194	0.354	0.084	-0.001	0.089	0.139	0.051	0.000	0.040	0.060	0.043
TSLs	0.318	0.331	0.380	0.395	0.085	0.111	0.159	0.132	0.017	0.046	0.068	0.065
JIVE	-0.079	0.356	6.520	0.059	-0.046	0.115	0.186	0.045	-0.008	0.046	0.069	0.045
LIML	-0.314	0.253	6.788	0.067	-0.020	0.103	0.177	0.048	-0.003	0.045	0.070	0.047
HLIM	-0.924	1.661	18.958	0.125	-0.193	0.169	0.432	0.028	-0.029	0.051	0.079	0.046
HFUL	6.581	1.460	291.007	0.104	-0.171	0.160	0.454	0.027	-0.027	0.049	0.078	0.046

Tables S.1 and S.2 generally show a good performance of estimators in both the ICM- and K-classes. At $\delta = 0.45$ in Table S.1 where identification is linear and weak, the MMD and IIV slightly over-reject while the TSLs, HLIM, and HFUL are size-distorted with substantial bias relative to the other estimators. In the related case of non-linear “semi-strong” identification in Table S.2, all estimators perform reasonably, although the TSLs still suffers size distortion at $\delta = 0.45$. In Tables S.3 and S.4, one observes improved performance of all estimators with increasing strength of non-parametric identification. No estimator controls size well in Table S.3 at $\delta = 0.1$ where non-parametric identification is linear and weak.

²Formally, the i 'th observation of the $2\dot{K}$, $\dot{K} = 6$ instrument vector is $[1, Z_i, W_{i1}, Z_i W_{i1}, \dots, W_{i, \dot{K}-1}, Z_i W_{i, \dot{K}-1}]$ where $W_{ik} = I(\Phi^{-1}((k-1)/\dot{K}) \leq Z_i \leq \Phi^{-1}(k/\dot{K}))$, $k = 1, \dots, \dot{K}$, and $\Phi^{-1}(\cdot)$ is the standard normal quantile function – see Antoine and Lavergne (2014, p. 64).

Table S.2: DGP_{2B} , $n = 250$

	$\delta = 0.45$				$\delta = 0.30$				$\delta = 0.15$			
	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.
MMD	-0.001	0.083	0.130	0.052	0.000	0.037	0.056	0.043	0.000	0.016	0.025	0.039
WMD	-0.011	0.074	0.118	0.041	-0.002	0.033	0.050	0.049	0.000	0.014	0.022	0.049
WMDF	-0.01	0.073	0.117	0.041	-0.001	0.032	0.050	0.049	0.000	0.014	0.022	0.049
DL	-0.003	0.079	0.124	0.050	0.000	0.035	0.054	0.044	0.000	0.015	0.024	0.040
ESC6	-0.004	0.079	0.125	0.051	0.000	0.035	0.054	0.047	0.000	0.015	0.024	0.040
IIV	0.000	0.074	0.113	0.048	0.000	0.033	0.049	0.047	0.000	0.014	0.022	0.048
TSLs	0.078	0.093	0.122	0.167	0.016	0.032	0.048	0.071	0.003	0.013	0.020	0.054
JIVE	-0.032	0.086	0.141	0.041	-0.005	0.032	0.051	0.045	-0.001	0.014	0.022	0.049
LIML	-0.017	0.077	0.134	0.040	-0.002	0.033	0.051	0.045	0.000	0.014	0.022	0.050
HLIM	-0.180	0.147	0.331	0.041	-0.025	0.039	0.061	0.059	-0.004	0.015	0.023	0.050
HFUL	-0.163	0.137	0.303	0.039	-0.023	0.038	0.060	0.056	-0.004	0.015	0.023	0.048

Table S.3: DGP_{3A} , $n = 250$

	$\delta = 0.1$				$\delta = 0.25$				$\delta = 0.5$			
	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.
MMD	0.252	0.360	0.802	0.204	0.004	0.155	0.270	0.082	0.001	0.081	0.126	0.053
WMD	0.424	0.424	7.439	0.105	-0.113	0.177	1.513	0.067	-0.011	0.090	0.142	0.053
WMDF	0.107	0.408	6.889	0.111	-0.065	0.176	0.480	0.068	-0.010	0.089	0.141	0.055
DL	0.026	0.426	28.89	0.182	0.037	0.162	0.896	0.072	-0.001	0.084	0.130	0.054
ESC6	0.252	0.347	0.618	0.179	-0.007	0.163	0.271	0.073	-0.002	0.084	0.130	0.055
IIV	0.190	0.404	2.143	0.195	0.003	0.173	0.286	0.081	0.000	0.088	0.136	0.060
TSLs	-0.255	0.401	5.934	0.094	-0.058	0.161	0.315	0.060	-0.012	0.081	0.127	0.047
JIVE	1.633	0.602	48.693	0.160	-0.275	0.185	3.012	0.045	-0.026	0.083	0.135	0.039
LIML	-0.255	0.401	5.934	0.094	-0.058	0.161	0.315	0.060	-0.012	0.081	0.127	0.047
HLIM	1.633	0.602	48.693	0.160	-0.275	0.185	3.012	0.045	-0.026	0.083	0.135	0.039
HFUL	0.060	0.394	5.869	0.102	-0.056	0.160	0.315	0.061	-0.011	0.080	0.127	0.048

Table S.4: DGP_{3B} , $n = 250$

	$\delta = 0.1$				$\delta = 0.5$				$\delta = 1.0$			
	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.
MMD	6.308	2.619	109.93	0.023	-0.099	0.732	9.107	0.041	-0.014	0.273	0.418	0.042
WMD	-1.180	2.567	96.641	0.017	-0.473	0.762	12.225	0.037	-0.021	0.276	0.430	0.037
WMDF	0.692	2.575	30.582	0.017	-0.461	0.762	10.319	0.037	-0.020	0.277	0.429	0.037
DL	-2.964	1.709	97.165	0.027	-0.062	0.398	0.661	0.039	-0.014	0.228	0.351	0.043
ESC6	-0.009	2.528	58.278	0.023	-0.178	0.625	1.864	0.049	-0.007	0.217	0.333	0.051
IIV	-6.919	2.581	181.213	0.018	-0.186	0.761	5.143	0.040	-0.015	0.275	0.424	0.039
TSLs	7.926	1.607	248.133	0.026	-0.048	0.344	0.555	0.040	-0.009	0.180	0.281	0.042
JIVE	3.388	1.681	57.376	0.089	-0.187	0.368	1.555	0.034	-0.021	0.183	0.285	0.040
LIML	7.926	1.607	248.133	0.026	-0.048	0.344	0.555	0.040	-0.009	0.180	0.281	0.042
HLIM	3.388	1.681	57.376	0.089	-0.187	0.368	1.555	0.034	-0.021	0.183	0.285	0.040
HFUL	0.038	1.699	42.708	0.026	-0.050	0.346	0.554	0.042	-0.005	0.179	0.279	0.044

S3.2. Bias and size control, $n = 500$

Table S.5 through Table S.11 present simulation results of DGP_{0A} through DGP_{3B} at sample size $n = 500$. One observes a general improvement in the performance of estimators in terms of reduced bias and better size control in settings where identification holds. A few points relative to the tables presented in the main text and the preceding subsection for $n = 250$ are worth noting. In Table S.7 at $\delta = 0.1$, the size distortion of the ESC6 remains even with the increase of the sample size to $n = 500$, although the MAD and RMSE decrease. Although reduced relative to the $n = 250$ case, the bias and

size distortion of the TSLS and HLIM in Table S.8 at $\delta = 0.45$ remain high which is clearly indicative of weak identification. The size distortion of the TSLS in Table S.9 at $\delta = 0.45$ is noticeable.

Table S.5: DGP_{0A} , $n = 500$

	$\delta = 0.1$				$\delta = 0.5$				$\delta = 1.0$			
	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.
<i>DGP_{0A}</i>												
MMD	-0.008	0.066	0.105	0.051	-0.001	0.030	0.046	0.057	0.000	0.021	0.032	0.054
WMD	-0.004	0.066	0.104	0.054	0.000	0.033	0.051	0.059	0.000	0.027	0.043	0.063
WMDF	-0.005	0.066	0.104	0.053	-0.001	0.033	0.051	0.059	0.000	0.027	0.043	0.063
DL	-0.023	0.087	0.141	0.031	-0.007	0.041	0.063	0.044	-0.004	0.030	0.046	0.049
ESC6	-0.015	0.073	0.116	0.034	-0.004	0.032	0.049	0.038	-0.002	0.023	0.034	0.040
IIV	-0.009	0.067	0.105	0.050	-0.003	0.033	0.052	0.061	-0.001	0.027	0.044	0.063
<i>DGP_{0B}</i>												
MMD	0.067	0.217	0.372	0.034	0.005	0.041	0.065	0.05	0.002	0.021	0.032	0.056
WMD	0.050	0.215	0.363	0.038	0.005	0.043	0.066	0.051	0.004	0.026	0.041	0.061
WMDF	0.051	0.216	0.365	0.038	0.005	0.043	0.066	0.051	0.004	0.026	0.041	0.061
DL	0.069	0.281	0.703	0.048	-0.004	0.056	0.085	0.053	-0.004	0.03	0.045	0.051
ESC6	0.085	0.244	0.449	0.034	0.006	0.044	0.069	0.040	0.002	0.022	0.034	0.040
IIV	0.061	0.220	0.371	0.035	0.007	0.043	0.067	0.051	0.005	0.026	0.041	0.062

Table S.6: DGP_{1A} , $n = 500$

	$\delta = 0.0$				$\delta = 0.25$				$\delta = 0.5$			
	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.
MMD	0.000	0.041	0.062	0.054	0.000	0.040	0.062	0.052	0.000	0.039	0.061	0.051
WMD	0.000	0.037	0.056	0.051	0.000	0.037	0.056	0.054	0.000	0.037	0.055	0.057
WMDF	0.000	0.037	0.056	0.051	0.000	0.037	0.056	0.054	0.000	0.037	0.055	0.057
DL	0.000	0.038	0.06	0.047	0.000	0.041	0.064	0.047	0.000	0.042	0.067	0.046
ESC6	0.002	0.036	0.054	0.058	0.002	0.036	0.054	0.056	0.002	0.035	0.053	0.055
IIV	0.002	0.038	0.058	0.056	0.002	0.038	0.058	0.054	0.002	0.038	0.057	0.057
TSLS	-1.208	0.781	15.955	0.005	0.697	0.331	19.261	0.017	-0.013	0.159	0.718	0.029
JIVE	0.052	0.487	26.628	0.240	-15.908	0.479	503.475	0.110	-0.217	0.189	3.894	0.027
LIML	-1.208	0.781	15.955	0.005	0.697	0.331	19.261	0.017	-0.013	0.159	0.718	0.029
HLIM	0.052	0.487	26.628	0.240	-15.908	0.479	503.475	0.110	-0.217	0.189	3.894	0.027
HFUL	9.676	1.013	211.364	0.022	-0.337	0.357	8.083	0.015	-0.085	0.160	0.780	0.023

S3.3. Size and power of the LC test

Define $\hat{\mathcal{E}}_i = D_i - \tilde{X}_i \hat{\eta}$ where $\hat{\eta}$ is an ICM estimator. From the literature on ICM specification testing, e.g., Su and Zheng (2017), the test statistic for the LC test based on an ICM kernel $K(\cdot, \cdot)$ is given by $T_n = nE_n[K(Z_i, Z_j)\hat{\mathcal{E}}_i\hat{\mathcal{E}}_j]$. T_n , like ICM specification test statistics in general, is not asymptotically pivotal under the null hypothesis and bootstrap methods are needed to compute the p-value. As the theoretical properties of ICM specification tests dating back to Bierens (1982) are already well established (see also Bierens and Ploberger (1997) and Su and Zheng (2017)), there is no need to pursue this path with respect to the proposed LC test. In this paper, the wild bootstrap, see, e.g., Escanciano (2006, p. 1042), with the Mammen (1993) two-point distribution for auxiliary variables is used to conduct all LC and specification tests.

Table S.7: DGP_{1B} , $n = 500$

	$\delta = 0.1$				$\delta = 0.5$				$\delta = 1.0$			
	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.
MMD	0.260	0.133	6.532	0.074	0.015	0.063	0.112	0.048	0.008	0.045	0.076	0.048
WMD	0.002	0.053	0.080	0.044	0.001	0.023	0.034	0.051	0.001	0.016	0.024	0.051
WMDF	0.003	0.053	0.079	0.044	0.001	0.023	0.034	0.051	0.001	0.016	0.024	0.052
DL	0.105	0.129	2.518	0.072	-0.256	0.060	8.186	0.047	0.006	0.043	0.086	0.045
ESC6	0.064	0.104	0.160	0.097	0.016	0.040	0.064	0.071	0.008	0.028	0.044	0.063
IIV	0.025	0.063	0.092	0.065	0.006	0.028	0.040	0.054	0.004	0.019	0.028	0.053
TSLs	0.062	0.269	10.041	0.000	-18.664	0.226	561.943	0.000	-1.602	0.163	44.187	0.001
JIVE	0.070	0.121	3.876	0.066	-0.190	0.103	4.942	0.048	0.225	0.069	4.164	0.021
LIML	0.062	0.269	10.041	0.000	-18.664	0.226	561.943	0.000	-1.602	0.163	44.187	0.001
HLIM	0.070	0.121	3.876	0.066	-0.190	0.103	4.942	0.048	0.225	0.069	4.164	0.021
HFUL	-0.343	0.140	13.319	0.029	1.583	0.113	37.819	0.026	-0.004	0.075	0.966	0.010

Table S.8: DGP_{2A} , $n = 500$

	$\delta = 0.45$				$\delta = 0.30$				$\delta = 0.15$			
	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.
MMD	-0.017	0.193	0.418	0.067	-0.003	0.078	0.125	0.049	-0.001	0.031	0.049	0.039
WMD	-0.047	0.204	1.67	0.054	-0.011	0.079	0.124	0.048	-0.001	0.031	0.047	0.044
WMDF	0.126	0.202	6.750	0.055	-0.011	0.078	0.123	0.048	-0.001	0.031	0.047	0.044
DL	-0.132	0.190	2.981	0.068	-0.005	0.077	0.122	0.047	-0.001	0.031	0.047	0.04
ESC6	-0.026	0.184	0.347	0.066	-0.005	0.077	0.122	0.048	-0.001	0.031	0.047	0.041
IIV	-0.004	0.196	0.358	0.075	0.000	0.077	0.120	0.052	0.000	0.031	0.047	0.045
TSLs	0.306	0.315	0.373	0.365	0.065	0.095	0.138	0.105	0.010	0.035	0.053	0.050
JIVE	1.241	0.327	64.119	0.046	-0.036	0.091	0.157	0.032	-0.006	0.035	0.054	0.038
LIML	-0.189	0.234	2.220	0.053	-0.016	0.088	0.146	0.035	-0.003	0.035	0.054	0.040
HLIM	0.357	1.556	27.494	0.108	-0.135	0.130	0.252	0.034	-0.018	0.036	0.058	0.047
HFUL	-0.217	1.300	27.413	0.082	-0.123	0.122	0.238	0.035	-0.017	0.036	0.057	0.046

Table S.9: DGP_{2B} , $n = 500$

	$\delta = 0.45$				$\delta = 0.30$				$\delta = 0.15$			
	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.
MMD	-0.001	0.079	0.124	0.049	0.000	0.031	0.049	0.040	0.000	0.012	0.019	0.043
WMD	-0.009	0.073	0.111	0.044	-0.001	0.029	0.043	0.039	0.000	0.011	0.017	0.038
WMDF	-0.009	0.073	0.111	0.047	-0.001	0.029	0.043	0.039	0.000	0.011	0.017	0.038
DL	-0.004	0.076	0.119	0.048	-0.001	0.030	0.047	0.041	0.000	0.012	0.018	0.044
ESC6	-0.004	0.077	0.120	0.048	-0.001	0.030	0.047	0.040	0.000	0.012	0.018	0.045
IIV	0.001	0.072	0.108	0.053	0.001	0.029	0.042	0.042	0.000	0.011	0.017	0.039
TSLs	0.076	0.087	0.118	0.164	0.013	0.029	0.041	0.076	0.002	0.011	0.016	0.044
JIVE	-0.030	0.083	0.135	0.040	-0.003	0.030	0.045	0.043	0.000	0.011	0.017	0.044
LIML	-0.013	0.072	0.122	0.031	-0.001	0.028	0.044	0.038	0.000	0.011	0.017	0.041
HLIM	-0.148	0.130	0.253	0.039	-0.017	0.032	0.049	0.054	-0.002	0.012	0.017	0.044
HFUL	-0.134	0.120	0.236	0.035	-0.016	0.032	0.049	0.053	-0.002	0.012	0.017	0.043

For simplicity, only the ICM specification test of Su and Zheng (2017) is employed for the LC test proposed in this paper (MMD LC).³ The LC test is conducted under three DGPs, namely, DGP_{0A} , DGP_{1B} , and DGP_{3B} .⁴ $\delta = 0.0$ in each DGP corresponds to the null hypothesis of a lack of non-parametric identification. 499 wild bootstrap samples and 1000 Monte Carlo simulations are used for the LC test. For comparison, the conventional first-stage IV F -test (IV FS) is considered for DGP_{1B} and DGP_{3B} where it is feasible.

³The Su and Zheng (2017) specification test and the MMD are based on the same kernel $K(Z, Z^\dagger) = -\|Z - Z^\dagger\|$.

⁴ DGP_{3B} is modified with a second covariate $\tilde{X} = Z_2$ for this subsection.

Table S.10: DGP_{3A} , $n = 500$

	$\delta = 0.1$				$\delta = 0.25$				$\delta = 0.5$			
	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.
MMD	0.08	0.263	1.632	0.137	0.002	0.112	0.181	0.063	0.001	0.058	0.088	0.054
WMD	0.529	0.313	12.827	0.083	-0.022	0.123	0.212	0.055	-0.004	0.063	0.096	0.044
WMDF	-0.019	0.311	1.953	0.085	-0.021	0.122	0.211	0.055	-0.004	0.063	0.096	0.044
DL	-3.220	0.288	98.206	0.129	-0.002	0.117	0.188	0.064	0.000	0.059	0.091	0.052
ESC6	0.105	0.255	0.475	0.118	-0.004	0.113	0.188	0.062	-0.001	0.059	0.091	0.052
IIV	0.173	0.290	1.670	0.137	0.002	0.121	0.195	0.066	0.001	0.063	0.095	0.048
TSL5	-0.304	0.270	3.522	0.074	-0.027	0.109	0.193	0.054	-0.006	0.055	0.087	0.041
JIVE	6.190	0.375	144.399	0.094	-0.061	0.117	0.225	0.036	-0.013	0.056	0.090	0.037
LIML	-0.304	0.270	3.522	0.074	-0.027	0.109	0.193	0.054	-0.006	0.055	0.087	0.041
HLIM	6.190	0.375	144.399	0.094	-0.061	0.117	0.225	0.036	-0.013	0.056	0.090	0.037
HFUL	-0.407	0.268	6.115	0.075	-0.027	0.110	0.193	0.053	-0.006	0.055	0.087	0.040

Table S.11: DGP_{3B} , $n = 500$

	$\delta = 0.1$				$\delta = 0.5$				$\delta = 1.0$			
	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.
MMD	4.790	2.569	73.194	0.016	-2.209	0.530	59.136	0.041	-0.015	0.195	0.289	0.043
WMD	1.128	2.643	47.062	0.015	-0.325	0.552	2.530	0.037	-0.016	0.196	0.293	0.046
WMDF	1.610	2.641	36.591	0.015	-0.322	0.552	2.507	0.037	-0.016	0.196	0.292	0.046
DL	-0.701	1.279	22.195	0.034	-0.025	0.280	0.436	0.037	-0.005	0.161	0.239	0.041
ESC6	10.606	2.636	218.732	0.015	-0.153	0.439	1.238	0.048	-0.007	0.158	0.231	0.048
IIV	-0.058	2.644	29.036	0.015	-0.298	0.551	2.374	0.041	-0.013	0.195	0.291	0.046
TSL5	-1.235	1.185	31.307	0.034	-0.027	0.234	0.379	0.033	-0.006	0.125	0.195	0.043
JIVE	0.615	1.515	18.730	0.100	-0.064	0.237	0.407	0.032	-0.012	0.126	0.196	0.044
LIML	-1.235	1.185	31.307	0.034	-0.027	0.234	0.379	0.033	-0.006	0.125	0.195	0.043
HLIM	0.615	1.515	18.730	0.100	-0.064	0.237	0.407	0.032	-0.012	0.126	0.196	0.044
HFUL	-0.880	1.235	33.766	0.032	-0.026	0.228	0.377	0.033	-0.004	0.125	0.194	0.041

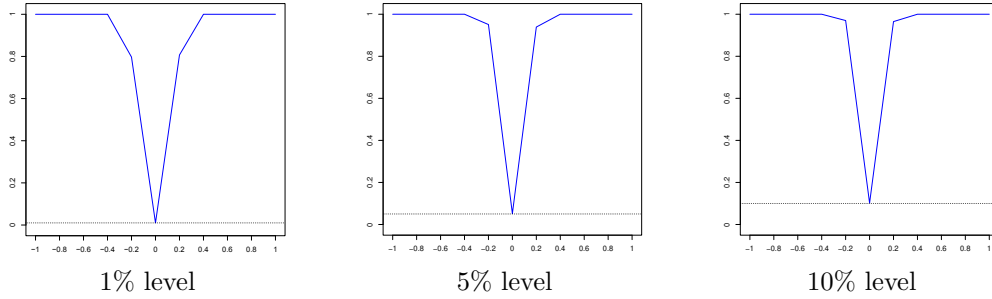


Figure S.4: Power Curve for the LC test - DGP_{0A}

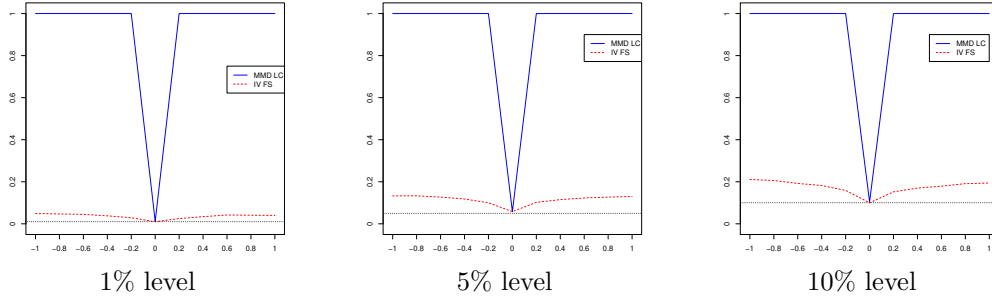


Figure S.8: Power Curve for the LC test - DGP_{1B}

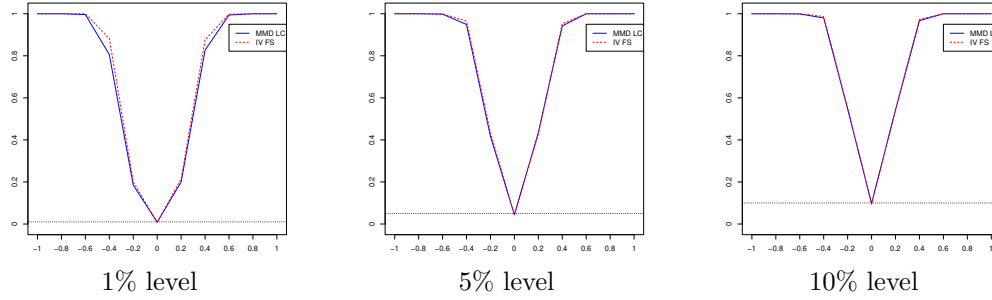


Figure S.12: Power Curve for the LC test - DGP_{3B}

Figures S.4, S.8 and S.12 plot the power curves of the LC test corresponding to DGP_{0A} , DGP_{1B} , and DGP_{3B} , respectively under the 1%, 5%, and 10% nominal levels. The dotted black horizontal line corresponds to the nominal level of the test. One observes from all nine power curves that the LC test controls size meaningfully and has non-trivial power against the alternative at all nominal levels. In Figure S.8, the standard first-stage IV relevance test has very low power as expected. The power curves of the LC and the standard first-stage IV relevance test in Figure S.12 are quite indistinguishable; both have good size control and comparable power. In sum, this short simulation exercise demonstrates the reliability of the proposed LC test.

S3.4. Sensitivity Analysis: Outliers

This subsection examines the sensitivity of ICM estimators to outliers in U . To this end, the following DGP is specified:

$$DGP_5 : \begin{cases} Y = \alpha_o + \beta_o D + U, & U = (1 - \Upsilon)U_N + \Upsilon U_C \\ D = Z + V \end{cases}$$

where $\Upsilon \sim \text{Ber}(\delta)$, $U_N \sim \mathcal{N}(0, 1)$, $\rho = \text{cov}[U_N, V] = 0.5$, and $U_C \sim \text{Cauchy}(0, 1)$. $\delta \in (0, 1)$ controls the degree of contamination of the error U with outliers.

Table S.12: DGP_5 , $n = 250$

	$\delta = 0.1$				$\delta = 0.25$				$\delta = 0.5$			
	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.
MMD	0.079	0.111	2.606	0.028	-0.296	0.228	14.907	0.021	-0.712	0.469	12.967	0.014
WMD	0.073	0.121	2.818	0.030	-0.246	0.255	17.178	0.017	-0.817	0.516	15.890	0.012
WMD _F	0.073	0.120	2.816	0.031	-0.246	0.255	17.166	0.017	-0.817	0.515	15.884	0.012
DL	0.082	0.115	2.556	0.032	-0.239	0.238	15.711	0.021	-0.736	0.501	13.738	0.012
ESC6	0.082	0.115	2.555	0.032	-0.230	0.236	15.940	0.021	-0.738	0.504	13.831	0.012
IIV	0.074	0.118	2.801	0.032	-0.244	0.253	17.106	0.017	-0.814	0.513	15.848	0.012
TSLs	0.078	0.106	2.422	0.030	-0.335	0.208	12.851	0.016	-0.629	0.437	10.706	0.016
JIVE	0.077	0.104	2.448	0.029	-0.339	0.207	13.001	0.017	-0.637	0.443	10.859	0.016
LIML	0.078	0.106	2.422	0.030	-0.335	0.208	12.851	0.016	-0.629	0.437	10.706	0.016
HLIM	0.077	0.104	2.448	0.029	-0.339	0.207	13.001	0.017	-0.637	0.443	10.859	0.016
HFUL	0.078	0.105	2.436	0.030	-0.335	0.207	12.924	0.017	-0.635	0.441	10.839	0.016

Tables S.12 and S.13 report results on the sensitivity analyses. As the presence of

Table S.13: DGP_5 , $n = 500$

	$\delta = 0.1$				$\delta = 0.25$				$\delta = 0.5$			
	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.
MMD	0.118	0.088	4.482	0.022	-0.559	0.202	24.128	0.022	-9.909	0.450	275.636	0.027
WMD	0.113	0.102	4.975	0.024	-0.716	0.220	30.342	0.022	-9.475	0.495	244.043	0.027
WMD ^F	0.113	0.102	4.975	0.024	-0.716	0.220	30.340	0.022	-9.474	0.495	244.010	0.027
DL	0.103	0.096	4.687	0.024	-0.710	0.212	28.508	0.024	-9.368	0.460	256.001	0.027
ESC6	0.103	0.097	4.690	0.023	-0.710	0.212	28.328	0.023	-9.287	0.462	253.658	0.027
IIV	0.114	0.102	4.966	0.025	-0.716	0.219	30.317	0.021	-9.398	0.494	242.378	0.026
TSLs	0.113	0.083	4.030	0.020	-0.426	0.183	18.452	0.021	-10.738	0.426	322.376	0.021
JIVE	0.113	0.083	4.048	0.019	-0.429	0.184	18.559	0.021	-10.831	0.428	324.110	0.021
LIML	0.113	0.083	4.030	0.020	-0.426	0.183	18.452	0.021	-10.738	0.426	322.376	0.021
HLIM	0.113	0.083	4.048	0.019	-0.429	0.184	18.559	0.021	-10.831	0.428	324.110	0.021
HFUL	0.114	0.083	4.044	0.020	-0.429	0.183	18.545	0.021	-10.792	0.427	323.039	0.021

outliers from the Cauchy distribution violates the dominance condition in Assumption 3.2(b), estimators of both the ICM- and K-classes are not expected to be consistent. The exercise, nonetheless, allows one to glean the performance of the estimators in the presence of contamination from a fat-tailed distribution, e.g., the Cauchy. One observes from both Tables S.12 and S.13 that MB, and RMSE deteriorate in sample size as these measures are less robust to outliers. The MAD, however, appears to improve in sample size but this is simply an artefact of the robustness of the MAD measure. The rejection rates suggest all estimators under-reject at the 5% nominal level. The above discussion confirms that in the presence of outliers in U , no estimator is relatively more robust.

S3.5. Do bounded one-to-one mappings alleviate the almost constant kernel problem?

In order to check whether one-to-one mappings alleviate the almost constant kernel problem in ICM estimators, simulations are presented in Table S.14 where Z is replaced with element-wise $\text{atan}(Z)$. A comparison of results in Table S.14 to those in Table 4 shows there are virtually no differences. This simply answers the question that the one-to-one mapping $\text{atan}(\cdot)$ does not alleviate the constant kernel problem. This is in large part because bounded one-to-one mappings cannot deal with the problem attributable to the multiplicity of instruments as suggested by the characterisation in Table 1 although they reduce the scale of instruments.

S3.6. Power under Alternatives

To alleviate concerns that the good size control of the MMD and ESC6 in Table 4 are not at the expense of power under alternatives, simulations which summarise the power of t -tests based on ICM estimators are presented in Figure S.13 with $n = 500$ and $p_z = 18$. $\beta - \beta_o \neq 0$ is used to examine the empirical power of the 5% t -test. Observe that the MMD and ESC6 perform reasonably. Their power curves are quite indistinguishable. Save the WMD, which controls size and has some power under the alternative, the power curves of the other estimators with multiplicative kernels show that they neither control size nor have power under the alternative.

Table S.14: DGP_4

	$p_z = 8$				$p_z = 18$				$p_z = 32$			
	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.	MB	MAD	RMSE	Rej.
$n = 250$												
MMD	0.007	0.034	0.049	0.061	0.014	0.031	0.047	0.072	0.024	0.036	0.051	0.122
WMD	-0.001	0.038	0.058	0.059	-0.003	0.062	0.098	0.056	0.134	0.127	0.161	0.500
WMDF	0.142	0.143	0.147	0.967	0.163	0.162	0.167	0.997	0.161	0.161	0.165	0.995
DL	0.003	0.043	0.066	0.064	0.159	0.165	0.168	0.851	0.161	0.161	0.165	0.995
ESC6	0.007	0.034	0.049	0.062	0.013	0.032	0.047	0.070	0.022	0.035	0.051	0.115
IIV	0.088	0.088	0.098	0.583	0.162	0.162	0.166	0.996	0.161	0.161	0.165	0.995
$n = 500$												
MMD	0.003	0.023	0.034	0.060	0.008	0.023	0.033	0.065	0.012	0.023	0.033	0.079
WMD	-0.001	0.026	0.040	0.065	-0.001	0.037	0.057	0.05	0.127	0.121	0.145	0.703
WMDF	0.094	0.093	0.098	0.905	0.164	0.164	0.166	1.000	0.161	0.161	0.163	1.000
DL	0.001	0.030	0.045	0.055	0.125	0.131	0.140	0.640	0.161	0.161	0.163	1.000
ESC6	0.003	0.022	0.034	0.061	0.008	0.023	0.033	0.064	0.012	0.022	0.033	0.076
IIV	0.057	0.056	0.066	0.445	0.163	0.163	0.165	1.000	0.161	0.161	0.163	1.000
$n = 1000$												
MMD	0.001	0.016	0.023	0.041	0.004	0.015	0.022	0.049	0.008	0.016	0.023	0.060
WMD	0.000	0.018	0.026	0.042	0.002	0.024	0.034	0.046	0.123	0.12	0.129	0.906
WMDF	0.040	0.039	0.045	0.396	0.164	0.163	0.165	1.000	0.161	0.161	0.162	1.000
DL	0.000	0.020	0.030	0.042	0.064	0.066	0.087	0.294	0.162	0.162	0.163	1.000
ESC6	0.001	0.016	0.023	0.042	0.004	0.015	0.022	0.047	0.007	0.016	0.023	0.056
IIV	0.035	0.035	0.042	0.295	0.162	0.161	0.163	1.000	0.161	0.161	0.162	1.000

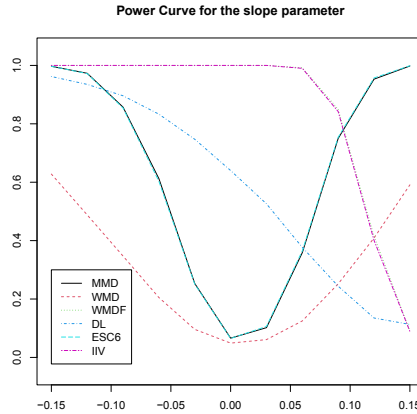


Figure S.13

S3.7. Computational Time of ICM Estimators

As the computation of an ICM estimator costs at least $O(n^2)$ steps, a crucial consideration for their practical usefulness is their computational cost. This subsection explores this question by providing running time as a function of the sample size n and the number of instruments p_z . Running times are reported for DGP_4 . For each pair $(n, p_z) \in \{250, 500, 1000\} \times \{8, 18, 32\}$, Table S.15 report the mean running time (MT) in seconds, the standard deviation (SDT) of the running time, the median computational time gain factor of the MMD (MGF), and the inter-quartile range of the computational gain factor (IQR). The computational gain factor of the MMD with respect to an estimator is computed as its running time divided by that of the MMD. In order to conduct

simulations in real time, the ESC6 and DL kernels are computed in C (Kernighan and Ritchie, 2006), whereas all else is computed in R (R Core Team, 2022). This implies computing the DL and ESC6 would incur considerably much higher computational cost (especially ESC6) if their kernels were implemented in R.⁵

Table S.15: Running Time DGP_4

	$p_z = 8$			$p_z = 18$			$p_z = 32$					
	MT	SDT	MGF	IQR	MT	SDT	MGF	IQR	MT	SDT	MGF	IQR
$n = 250$												
MMD	0.011	0.004	1.000	0.000	0.012	0.003	1.000	0.000	0.014	0.004	1.000	0.000
WMD	0.014	0.005	1.250	0.545	0.015	0.004	1.222	0.500	0.017	0.004	1.273	0.446
WMDF	0.013	0.005	1.182	0.518	0.015	0.004	1.222	0.500	0.016	0.004	1.200	0.455
DL	0.138	0.016	13.04	4.792	0.146	0.017	12.57	4.059	0.149	0.018	11.38	3.773
ESC6	1.577	0.127	151.8	52.44	2.981	0.322	257.6	81.89	4.883	0.499	372.3	114.6
IIV	0.017	0.006	1.500	0.657	0.022	0.007	1.778	0.688	0.033	0.008	2.462	0.933
$n = 500$												
MMD	0.023	0.006	1.000	0.000	0.027	0.006	1.000	0.000	0.028	0.006	1.000	0.000
WMD	0.029	0.006	1.239	0.419	0.034	0.010	1.241	0.397	0.035	0.006	1.229	0.364
WMDF	0.027	0.006	1.190	0.414	0.033	0.009	1.185	0.424	0.033	0.006	1.156	0.393
DL	1.137	0.106	49.19	15.90	1.169	0.118	43.62	13.75	1.086	0.058	38.58	11.88
ESC6	13.20	1.020	575.7	200.0	24.35	2.081	911.4	272.0	36.85	1.374	1314.0	383.5
IIV	0.026	0.006	1.154	0.379	0.035	0.012	1.263	0.393	0.044	0.009	1.542	0.412
$n = 1000$												
MMD	0.054	0.012	1.000	0.000	0.065	0.015	1.000	0.000	0.082	0.028	1.000	0.000
WMD	0.069	0.009	1.354	0.295	0.081	0.013	1.310	0.245	0.099	0.031	1.258	0.199
WMDF	0.068	0.008	1.386	0.449	0.079	0.011	1.326	0.364	0.097	0.026	1.262	0.301
DL	7.691	0.397	157.9	51.55	7.864	0.685	133.5	39.56	8.313	1.270	111.4	25.50
ESC6	84.21	3.246	1720.7	550.3	151.2	13.27	2555.3	709.6	317.4	58.88	4196.7	941.9
IIV	0.058	0.006	1.133	0.279	0.072	0.012	1.149	0.250	0.100	0.027	1.239	0.254

Particular emphasis is put on the computational cost of MMD and ESC6 as these are the only ICM estimators that appear viable under DGP_4 especially at $p_z \in \{18, 32\}$ – see Table 4. A clearly discernible pattern from the MT and MGF is that the MMD is the most computationally efficient, albeit only marginally so with respect to the WMD, WMDF, and IIV. The marginal computational gain is not surprising as the WMD, WMDF, and IIV kernels only involve simple non-linear transformations of the MMD kernel. With respect to the DL and especially ESC6, one observes a tremendous difference in computational time. Observe from the mean computational times (MT) of the MMD, WMD, WMDF, and the IIV that computations for the given sample sizes n and number of instruments p_z are reasonably fast. The mean computational time of the DL seems much but not excessive. The mean running times of the ESC6 suggest it is very computationally costly. Across all dimensions $p_z \in \{8, 18, 32\}$, doubling the sample size from 250 to 500 seems to augment the mean running time of the ESC6 by a factor of at least 7.5 while increasing n from 500 to 1000 seems to augment the mean running time by a factor of at least 6.

The median computational time gain factor (MGF) of the MMD is perhaps the most informative on the relative computational gain of the MMD vis-à-vis ESC6. Observe that although the ESC6 kernel is computed in C, the ESC6 is still computationally costly by

⁵All computations are conducted on a MacBook Pro computer with 2.8 GHz, Quad-Core Intel Core i7, and 16 GB.

a considerable factor relative to the MMD. For example, at sample size $n = 1000$ and $p_z = 18$ and the same level of performance in terms of MB, MAD, and RMSE (see Table 4), the MMD is more than 2 500 times faster than the ESC6. By doubling the sample size, the computational time gain factor triples approximately across all three values of p_z considered. Also observe that the computational time gain factor increases with p_z . The foregoing results suggest that the ESC6, although well-performing in reducing bias and size control just like the MMD in the presence of multiple instruments, is computationally much more costly than the MMD.

REFERENCES

- Antoine, B. and P. Lavergne (2014). Conditional moment models under semi-strong identification. *Journal of Econometrics* 182, 59–69.
- Bierens, H. J. (1982). Consistent model specification tests. *Journal of Econometrics* 20, 105–134.
- Bierens, H. J. and W. Ploberger (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica* 65, 1129–1151.
- Escanciano, J. C. (2006). A consistent diagnostic test for regression models using projections. *Econometric Theory* 22, 1030–1051.
- Escanciano, J. C. (2018). A simple and robust estimator for linear regression models with strictly exogenous instruments. *Econometrics Journal* 21, 36–54.
- Hansen, B. E. (2021). *Econometrics*. University of Wisconsin, Dept. of Economics.
- Hoeffding, W. (1961). The strong law of large numbers for U-statistics. Working Paper, North Carolina State University.
- Johnson, N. L., S. Kotz, and N. Balakrishnan (1995). *Continuous univariate distributions, Volume 1* (2nd ed.). New York, NY: John Wiley & Sons.
- Kernighan, B. W. and D. M. Ritchie (2006). *The C programming language*.
- Mammen, Enno (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Annals of Statistics* 21, 255–285.
- Shao, X. and J. Zhang (2014). Martingale difference correlation and its use in high-dimensional variable screening. *Journal of the American Statistical Association* 109, 1302–1318.
- Su, L. and X. Zheng (2017). A martingale-difference-divergence-based test for specification. *Economics Letters* 156, 162–167.
- Székely, G. J., M. L. Rizzo, and N. K. Bakirov (2007). Measuring and testing dependence by correlation of distances. *Annals of Statistics* 35, 2769–2794.
- R Core Team (2022). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.