SUPPLEMENTARY MATERIALS FOR:
"RECONCILING MODEL-X AND DOUBLY ROBUST APPROACHES TO CONDITIONAL INDEPENDENCE TESTING"

BY ZIANG NIU[*,1,a], ABHINAV CHAKRABORTY[*,1,b],
OLIVER DUKES[2,d] AND EUGENE KATSEVICH[1,c]

[1]*Department of Statistics and Data Science, University of Pennsylvania ,* [a]*ziangniu@wharton.upenn.edu;*
[b]*abch@wharton.upenn.edu;* [c]*ekatsevi@wharton.upenn.edu*

[2]*Department of Applied Mathematics, Computer Science and Statistics, Ghent University,* [d]*oliver.dukes@ugent.be*

APPENDIX A: THE $\widehat{\text{dCRT}}$ WITH GCM NORMALIZATION

As an alternative to the $\widehat{\text{dCRT}}$, we consider the $\widehat{\text{ndCRT}}$. This procedure is based on a normalized statistic that coincides exactly with the GCM statistic:

$$T_n^{\widehat{\text{ndCRT}}}(X,Y,Z) \equiv \frac{1}{\widehat{S}_n^{\text{GCM}}} T_n^{\widehat{\text{dCRT}}}(X,Y,Z) \equiv T_n^{\text{GCM}}(X,Y,Z).$$

The only difference with GCM is that the critical value is given by conditional resampling rather than a normal quantile:

(S1)
$$\begin{aligned}
\phi_n^{\widehat{\text{ndCRT}}}(X,Y,Z) &\equiv \mathbb{1}(T_n^{\widehat{\text{ndCRT}}}(X,Y,Z) > \mathbb{Q}_{1-\alpha}[T_n^{\widehat{\text{ndCRT}}}(\widetilde{X},X,Y,Z) \mid X,Y,Z]) \\
&\equiv \mathbb{1}\left(T_n^{\widehat{\text{ndCRT}}}(X,Y,Z) > C_n^{\widehat{\text{ndCRT}}}(X,Y,Z)\right).
\end{aligned}$$

Here,

$$T_n^{\widehat{\text{ndCRT}}}(\widetilde{X},X,Y,Z) \equiv \frac{T_n^{\text{dCRT}}(\widetilde{X},X,Y,Z)}{S_n^{\text{GCM}}(\widetilde{X},X,Y,Z)},$$

where

$$(S_n^{\text{GCM}}(\widetilde{X},X,Y,Z))^2 \equiv \widehat{\text{Var}}\{(\widetilde{X}_i - \widehat{\mu}_{n,x}(Z_i))(Y_i - \widehat{\mu}_{n,y}(Z_i))\}$$

and $T_n^{\text{dCRT}}(\widetilde{X},X,Y,Z)$ is as defined in equation (5).

THEOREM 1. *Let $\mathcal{L}_n$ be a sequence of laws such that the nondegeneracy conditions* (NDG1) *and* (NDG2) *and the conditional Lyapunov condition* (Lyap-1) *hold. Then,*

(S2)
$$T_n^{\widehat{\text{ndCRT}}}(\widetilde{X},X,Y,Z) \mid X,Y,Z \xrightarrow{d,p} N(0,1)$$

and therefore

(S3)
$$C_n^{\widehat{\text{ndCRT}}}(X,Y,Z) \equiv \mathbb{Q}_{1-\alpha}\left[T_n^{\widehat{\text{ndCRT}}}(\widetilde{X},X,Y,Z) \mid X,Y,Z\right] \to z_{1-\alpha}$$

and the $\widehat{\text{ndCRT}}$ and GCM tests are equivalent:

(S4)
$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{L}_n}[\phi_n^{\widehat{\text{ndCRT}}}(X,Y,Z) = \phi_n^{\text{GCM}}(X,Y,Z)] = 1.$$

COROLLARY 1. *Let $\mathscr{R}_n$ be a sequence of regularity conditions such that for any sequence $\mathcal{L}_n \in \mathscr{R}_n$, we have the the nondegeneracy conditions* (NDG1) *and* (NDG2)*, the conditional Lyapunov condition* (Lyap-1)*, and the assumptions* (SP1) *and* (SP2)*. Then, the $\widehat{\text{ndCRT}}$ has asymptotic Type-I error control over $\mathscr{L}_n^0 \cap \mathscr{R}_n$ in the sense of the definition* (28)*.*

Comparing Corollary 1 to Corollary 3, we see that the $\widehat{\text{ndCRT}}$ controls Type-I error under weaker assumptions than required for the $\widehat{\text{dCRT}}$; in particular the variance of $\mathcal{L}_n(\boldsymbol{X}|\boldsymbol{Z})$ need not be estimated with any degree of accuracy.

APPENDIX B: CONDITIONAL CONVERGENCE RESULTS

The proofs of our theoretical results rely on the conditional counterparts of several standard convergence theorems. In this section, we state these conditional convergence theorems. We defer their proofs to Appendix G.

First we define a notion of conditional convergence in probability, analogous to our definition of conditional convergence in distribution (Definition 1).

DEFINITION 1. For each n, let W_n be a random variable and let $\mathcal{F}_n$ be a σ-algebra. Then, we say W_n converges in probability to a constant c conditionally on $\mathcal{F}_n$ if W_n converges in distribution to the delta mass at c conditionally on $\mathcal{F}_n$ (recall Definition 1). We denote this convergence by $W_n \mid \mathcal{F}_n \xrightarrow{p,p} c$. In symbols,

$$(\text{S5}) \qquad W_n \mid \mathcal{F}_n \xrightarrow{p,p} c \quad \text{if} \quad W_n \mid \mathcal{F}_n \xrightarrow{d,p} \delta_c.$$

Now we are ready to state the conditional convergence results.

B.1. Statements. For the sake of all results below, let $\mathcal{F}_n$ be a sequence of σ-algebras.

THEOREM 2 (Conditional Polya's theorem). *Let W_n be a sequence of random variables. If $W_n \mid \mathcal{F}_n \xrightarrow{d,p} W$ for some random variable W with continuous CDF, then*

$$(\text{S6}) \qquad \sup_{t \in \mathbb{R}} |\mathbb{P}[W_n \leq t \mid \mathcal{F}_n] - \mathbb{P}[W \leq t]| \xrightarrow{p} 0.$$

THEOREM 3 (Conditional Slutsky's theorem). *Let W_n be a sequence of random variables. Suppose a_n and b_n are sequences of random variables such that $a_n \xrightarrow{p} 1$ and $b_n \xrightarrow{p} 0$. If $W_n \mid \mathcal{F}_n \xrightarrow{d,p} W$ for some random variable W with continuous CDF, then*

$$(\text{S7}) \qquad a_n W_n + b_n \mid \mathcal{F}_n \xrightarrow{d,p} W.$$

THEOREM 4 (Conditional law of large numbers). *Let W_{in} be a triangular array of random variables, such that W_{in} are independent conditionally on $\mathcal{F}_n$ for each n. If for some $\delta > 0$ we have*

$$(\text{S8}) \qquad \frac{1}{n^{1+\delta}} \sum_{i=1}^{n} \mathbb{E}[|W_{in}|^{1+\delta} \mid \mathcal{F}_n] \xrightarrow{p} 0,$$

then

$$(\text{S9}) \qquad \frac{1}{n} \sum_{i=1}^{n} (W_{in} - \mathbb{E}[W_{in}|\mathcal{F}_n]) \mid \mathcal{F}_n \xrightarrow{p,p} 0.$$

The condition (S8) *is satisfied when*

$$(\text{S10}) \qquad \sup_{1 \leq i \leq n} \mathbb{E}[|W_{in}|^{1+\delta} \mid \mathcal{F}_n] = o_p(n^{\delta}).$$

As a corollary of Theorem 4, if we choose $\mathcal{F}_n = \{\varnothing, \Omega\}$, we are able to obtain the following version of the weak law of large numbers for triangular arrays.

COROLLARY 2 (Unconditional weak law of large numbers). *Let W_{in} be a triangular array of random variables, such that W_{in} are independent for each n. If for some $\delta > 0$ we have*

$$(S11) \qquad \frac{1}{n^{1+\delta}} \sum_{i=1}^{n} \mathbb{E}[|W_{in}|^{1+\delta}] \to 0,$$

then

$$(S12) \qquad \frac{1}{n} \sum_{i=1}^{n} (W_{in} - \mathbb{E}[W_{in}]) \overset{p}{\to} 0.$$

The condition (S11) *is satisfied when*

$$(S13) \qquad \sup_{1 \le i \le n} \mathbb{E}[|W_{in}|^{1+\delta}] = o(n^{\delta}).$$

THEOREM 5 (Conditional central limit theorem). *Let W_{in} be a triangular array of random variables, such that for each n, W_{in} are independent conditionally on $\mathcal{F}_n$. Define*

$$(S14) \qquad S_n^2 \equiv \sum_{i=1}^{n} \mathrm{Var}[W_{in} \mid \mathcal{F}_n],$$

and assume $0 < \mathrm{Var}[W_{in}|\mathcal{F}_n] < \infty$ almost surely for all $i = 1, \ldots, n$ and for all $n \in \mathbb{N}$. If for some $\delta > 0$ we have

$$(S15) \qquad \frac{1}{S_n^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}[|W_{in} - \mathbb{E}[W_{in}|\mathcal{F}_n]|^{2+\delta} \mid \mathcal{F}_n] \overset{p}{\to} 0,$$

then

$$(S16) \qquad \frac{1}{S_n} \sum_{i=1}^{n} (W_{in} - \mathbb{E}[W_{in} \mid \mathcal{F}_n]) \mid \mathcal{F}_n \overset{d,p}{\longrightarrow} N(0,1).$$

LEMMA 1 (Conditional convergence implies quantile convergence). *Let W_n be a sequence of random variables and $\alpha \in (0,1)$. If $W_n \mid \mathcal{F}_n \overset{d,p}{\longrightarrow} W$ for some random variable W whose CDF is continuous and strictly increasing at $\mathbb{Q}_\alpha[W]$, then*

$$(S17) \qquad \mathbb{Q}_\alpha[W_n \mid \mathcal{F}_n] \overset{p}{\to} \mathbb{Q}_\alpha[W].$$

B.2. Discussion. The above definitions and results on conditional convergence are not particularly surprising, and related results are present in the existing literature. Nevertheless, we have not found any of the above results stated in the literature in exactly this form. Here we discuss the relationships of our definitions and results with existing ones.

Notions of conditional convergence in probability and in distribution have been explicitly defined by Nowak and Zięba (2005). However, these notions require a single conditioning σ-algebra as well as almost sure convergences of conditional probabilities, whereas in Definitions 1 and 1 we allow the conditioning σ-algebra to change with n and for the conditional probabilities to converge in probability. Our Definition 1 can be viewed as formalizing the notion of conditional convergence in distribution implicitly used by Wang and Janson (2022). Related notions of conditional convergence in distribution allowing for changing conditioning σ-algebra are present implicitly in the works of Dedecker and Merlevede (2002) and Bulinski (2017), though these are based on the convergence of conditional characteristic functions as opposed to conditional cumulative distribution functions.

Turning to the convergence results themselves, we were not able to find conditional Polya's theorem (Theorem 2) in the literature. Conditional Slutsky's theorem (Theorem 3) is a generalization of Wang and Janson (2022, Lemma 5) to the case when a_n is not necessarily independent of $\mathcal{F}_n$ and $b_n \neq 0$. Versions of the conditional law of large numbers are given by Majerek, Nowak and Zieba (2005) and Prakasa Rao (2009), but these involve a single conditioning σ-algebra and do not allow for triangular arrays, unlike Theorem 4. Remarkably, we could not find even the unconditional triangular array law of large numbers (Corollary 2) in the literature; existing results either assume a second-moment condition or use truncation (Durrett, 2010, Theorems 2.2.4 and 2.2.6, respectively) instead of a $1 + \delta$ moment condition or are not applicable to triangular arrays (Shah and Peters, 2020, Lemma 19). As for central limit theorems, Grzenda and Zieba (2008); Prakasa Rao (2009); Yuan, Wei and Lei (2014) give non-triangular array versions of the conditional central limit theorem that require a single conditioning σ-algebra, unlike Theorem 5. Versions of the conditional central limit theorem appropriate for varying conditioning σ-algebras and triangular arrays are given by Dedecker and Merlevede (2002); Bulinski (2017), those these involve different notions of conditional convergence in distribution. Results similar to Theorem 5 for $\delta = 1$ are presented in a recent line of work on sample-splitting-based inference (Kim and Ramdas, 2024; Shekhar, Kim and Ramdas, 2022, 2023); these can be proved via the Berry-Esseen theorem. Finally, we note that our result that conditional convergence in distribution implies in-probability quantile convergence (Lemma 1) is a generalization of Wang and Janson (2022, Lemma 3) to general conditioning σ-algebras.

APPENDIX C: PROOFS FOR SECTION 2

C.1. Auxiliary lemmas.

LEMMA 2. *Let W_n be a sequence of nonnegative random variables and let $\mathcal{F}_n$ be a sequence of σ-algebras. If $\mathbb{E}[W_n \mid \mathcal{F}_n] \xrightarrow{p} 0$, then $W_n \xrightarrow{p} 0$.*

PROOF. For any $\epsilon > 0$, we have

$$\text{(S18)} \qquad \mathbb{P}[W_n \geq \epsilon] = \mathbb{P}[W_n \wedge \epsilon \geq \epsilon]$$

$$\text{(S19)} \qquad \leq \epsilon^{-1} \mathbb{E}[W_n \wedge \epsilon]$$

$$\text{(S20)} \qquad = \epsilon^{-1} \mathbb{E}[\mathbb{E}[W_n \wedge \epsilon \mid \mathcal{F}_n]]$$

$$\text{(S21)} \qquad \leq \epsilon^{-1} \mathbb{E}[\mathbb{E}[W_n \mid \mathcal{F}_n] \wedge \epsilon] \to 0,$$

where the last convergence is due to bounded convergence theorem and the assumption $\mathbb{E}[W_n \mid \mathcal{F}_n] \xrightarrow{p} 0$. □

LEMMA 3 (Asymptotic equivalence of tests). *Consider two hypothesis tests based on the same test statistic $T_n(X, Y, Z)$ but different critical values:*

$$\phi_n^1(X, Y, Z) \equiv \mathbb{1}(T_n(X, Y, Z) > C_n(X, Y, Z)); \quad \phi_n^2(X, Y, Z) \equiv \mathbb{1}(T_n(X, Y, Z) > z_{1-\alpha}).$$

If the critical value of the first converges in probability to that of the second:

$$\text{(S22)} \qquad C_n(X, Y, Z) \xrightarrow{p} z_{1-\alpha}$$

and the test statistic does not accumulate near the limiting critical value:

$$\text{(S23)} \qquad \lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P}_{\mathcal{L}_n}[|T_n(X, Y, Z) - z_{1-\alpha}| \leq \delta] = 0,$$

then the two tests are asymptotically equivalent:

$$\text{(S24)} \qquad \lim_{n \to \infty} \mathbb{P}_{\mathcal{L}_n}[\phi_n^1(X, Y, Z) = \phi_n^2(X, Y, Z)] = 1.$$

PROOF. Note that for any $\delta > 0$, we have

$$\mathbb{P}_{\mathcal{L}_n}\left[\phi_n^1(X,Y,Z) \neq \phi_n^2(X,Y,Z)\right]$$

$$= \mathbb{P}_{\mathcal{L}_n}\left[\min\left(z_{1-\alpha}, C_n\right) < T_n \leq \max\left(z_{1-\alpha}, C_n\right)\right]$$

(S25)
$$= \mathbb{P}_{\mathcal{L}_n}\left[\min\left(z_{1-\alpha}, C_n\right) < T_n \leq \max\left(z_{1-\alpha}, C_n\right), |C_n - z_{1-\alpha}| \leq \delta\right]$$

$$+ \mathbb{P}_{\mathcal{L}_n}\left[\min\left(z_{1-\alpha}, C_n\right) < T_n \leq \max\left(z_{1-\alpha}, C_n\right), |C_n - z_{1-\alpha}| > \delta\right]$$

$$\leq \mathbb{P}_{\mathcal{L}_n}\left[|T_n - z_{1-\alpha}| \leq \delta\right] + \mathbb{P}_{\mathcal{L}_n}\left[|C_n - z_{1-\alpha}| > \delta\right].$$

To justify the last step, suppose without loss of generality that $z_{1-\alpha} \leq C_n$. Then note that if $z_{1-\alpha} < T_n \leq C_n$ and $C_n - z_{1-\alpha} \leq \delta$ then

$$|T_n - z_{1-\alpha}| = T_n - z_{1-\alpha} \leq C_n - z_{1-\alpha} \leq \delta.$$

Taking a limsup on both sides in equation (S25) and using the assumed convergence (S22), we find that

$$\limsup_{n\to\infty} \mathbb{P}_{\mathcal{L}_n}\left[\phi_n^1(X,Y,Z) \neq \phi_n^2(X,Y,Z)\right]$$

$$\leq \limsup_{n\to\infty} \mathbb{P}_{\mathcal{L}_n}\left[|T_n(X,Y,Z) - z_{1-\alpha}| \leq \delta\right] + \limsup_{n\to\infty} \mathbb{P}_{\mathcal{L}_n}\left[|C_n(X,Y,Z) - z_{1-\alpha}| > \delta\right]$$

$$= \limsup_{n\to\infty} \mathbb{P}_{\mathcal{L}_n}\left[|T_n(X,Y,Z) - z_{1-\alpha}| \leq \delta\right].$$

Letting $\delta \to 0$ and using our assumption (S23), we arrive at the claimed asymptotic equivalence. This completes the proof. $\qquad\square$

C.2. Proofs of main results.

PROOFS OF THEOREM 1 AND COROLLARY 1. We prove instead the stronger Theorem 6 and Corollary 3 below. $\qquad\square$

THEOREM 6. *Let $\mathcal{L}_n$ be a sequence of laws and $\widehat{\mathcal{L}}_n$ be a sequence of estimates. Suppose either of the following two sets of assumptions is satisfied:*

1. *The nondegeneracy conditions (NDG1) and (NDG2) and the conditional Lyapunov condition (Lyap-1) hold.*
2. *The assumptions of Theorem 2 hold.*

Then, the normalized $T_n^{\widehat{\mathrm{dCRT}}}(\widetilde{X}, X, Y, Z)$ converges conditionally to a normal distribution (12) and therefore the $\widehat{\mathrm{dCRT}}$ critical value $C_n^{\widehat{\mathrm{dCRT}}}(X,Y,Z)$ converges to $z_{1-\alpha}$ (13).

PROOF. It suffices to prove the conditional convergence in distribution (12), as the convergence of the critical value (13) follows from Lemma 1 because the normal distribution has continuous and strictly increasing CDF. We prove the conditional convergence (12) for each of the two sets of assumptions.

Assumption 1. We proceed by applying the conditional CLT (Theorem 5) with

(S26)
$$W_{in} \equiv \frac{1}{\sqrt{n}}(\widetilde{X}_i - \widehat{\mu}_{n,x}(Z_i))(Y_i - \widehat{\mu}_{n,y}(Z_i))$$

and $\mathcal{F}_n \equiv \sigma(X,Y,Z)$. To verify the assumptions of the conditional CLT, note first that W_{in} are independent conditionally on $\mathcal{F}_n$ by construction and satisfy $0 < \mathrm{Var}[W_{in} \mid \mathcal{F}_n] < \infty$ by the nondegeneracy assumption (NDG2). Next, recalling definition (9), we have

$$S_n^2 \equiv \sum_{i=1}^n \mathrm{Var}_{\widehat{\mathcal{L}}_n}[W_{in} \mid X,Y,Z] \equiv (\widehat{S}_n^{\widehat{\mathrm{dCRT}}})^2,$$

so that

$$\frac{1}{S_n^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}_{\widehat{\mathcal{L}}_n}[|W_{in}|^{2+\delta} \mid X, Y, Z]$$

$$= \frac{1}{(\widehat{S}_n^{\widehat{\mathrm{dCRT}}})^{2+\delta}} \cdot \frac{1}{n^{1+\delta/2}} \sum_{i=1}^{n} |Y_i - \widehat{\mu}_{n,y}(Z_i)|^{2+\delta} \mathbb{E}_{\widehat{\mathcal{L}}_n}\left[\left|\widetilde{X}_i - \widehat{\mu}_{n,x}(Z_i)\right|^{2+\delta} \mid X, Z\right].$$

This quantity converges to zero in probability due to the nondegeneracy condition (NDG1) and the Lyapunov condition (Lyap-1). Hence, the conditional CLT gives the desired conditional convergence (12).

Assumption 2. We begin by decomposing $T_n^{\widehat{\mathrm{dCRT}}}(\widetilde{X}, X, Y, Z)$:

(S27)
$$T_n^{\widehat{\mathrm{dCRT}}}(\widetilde{X}, X, Y, Z) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\widetilde{X}_i - \widehat{\mu}_{n,x}(Z_i))(Y_i - \mu_{n,y}(Z_i))$$

(S28)
$$- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\widetilde{X}_i - \widehat{\mu}_{n,x}(Z_i))(\widehat{\mu}_{n,y}(Z_i) - \mu_{n,y}(Z_i))$$

(S29)
$$\equiv I_n - J_n.$$

We claim that $J_n \xrightarrow{p} 0$. Indeed, from $\mathbb{E}[J_n \mid X, Y, Z] = 0$ and the assumption $\widehat{E}'_{n,y} \xrightarrow{p} 0$ (SP1') it follows

$$\mathbb{E}[J_n^2 \mid X, Y, Z] = \mathrm{Var}[J_n \mid X, Y, Z] \equiv (\widehat{E}'_{n,y})^2 \xrightarrow{p} 0.$$

Hence by Lemma 2 we have $J_n^2 \xrightarrow{p} 0$, so that $J_n \xrightarrow{p} 0$, as claimed. Next, we claim that an appropriately rescaled I_n converges conditionally to $N(0,1)$. To this end, we apply the conditional CLT (Theorem 5) with

(S30)
$$W_{in} \equiv \frac{1}{\sqrt{n}}(\widetilde{X}_i - \widehat{\mu}_{n,x}(Z_i))(Y_i - \mu_{n,y}(Z_i))$$

and $\mathcal{F}_n \equiv \sigma(X, Y, Z)$. To verify the assumptions of the conditional CLT, note first that W_{in} are independent conditionally on $\mathcal{F}_n$ by construction and satisfy $0 < \mathrm{Var}[W_{in} \mid \mathcal{F}_n] < \infty$ by the nondegeneracy assumption (NDG2). Next, observe that

$$S_n^2 \equiv \sum_{i=1}^{n} \mathrm{Var}[W_{in} \mid X, Y, Z] = \frac{1}{n} \sum_{i=1}^{n} \mathrm{Var}_{\widehat{\mathcal{L}}_n}[X_i|Z_i](Y_i - \mu_{n,y}(Z_i))^2 \equiv (S_n^{\widehat{\mathrm{dCRT}}})^2,$$

so that

$$\frac{1}{S_n^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}[|W_{in}|^{2+\delta} \mid X, Y, Z]$$

$$= \frac{1}{(S_n^{\widehat{\mathrm{dCRT}}})^{2+\delta}} \cdot \frac{1}{n^{1+\delta/2}} \sum_{i=1}^{n} |Y_i - \mu_{n,y}(Z_i)|^{2+\delta} \mathbb{E}_{\widehat{\mathcal{L}}_n}[|\widetilde{X}_i - \widehat{\mu}_{n,x}(Z_i)|^{2+\delta} \mid X, Z].$$

Since the first factor is stochastically bounded (conclusion (S49) from Lemma 7), it suffices to show that the second factor converges to zero in probability. To this end, by Lemma 2 it

suffices to note that $\mathcal{L}_n \in \mathscr{L}_n^0$ and the Lyapunov assumption (Lyap-2) give

$$
(\text{S31}) \quad
\begin{aligned}
&\mathbb{E}\left[\frac{1}{n^{1+\delta/2}}\sum_{i=1}^n |Y_i - \mu_{n,y}(Z_i)|^{2+\delta}\mathbb{E}_{\widehat{\mathcal{L}}_n}[|\widetilde{X}_i - \widehat{\mu}_{n,x}(Z_i)|^{2+\delta} \mid X, Z]\,\bigg|\, X, Z\right] \\
&\quad = \frac{1}{n^{1+\delta/2}}\sum_{i=1}^n \mathbb{E}_{\mathcal{L}_n}[|Y_i - \mu_{n,y}(Z_i)|^{2+\delta}|Z_i]\mathbb{E}_{\widehat{\mathcal{L}}_n}[|\widetilde{X}_i - \widehat{\mu}_{n,x}(Z_i)|^{2+\delta} \mid X, Z] \\
&\quad \xrightarrow{p} 0.
\end{aligned}
$$

Therefore, we may apply the conditional CLT to obtain that

$$
\frac{1}{S_n^{\widehat{\text{dCRT}}}} I_n \mid X, Y, Z \xrightarrow{d,p} N(0,1).
$$

Furthermore, equation (S48) from Lemma 7 gives $S_n^{\widehat{\text{dCRT}}}/\widehat{S}_n^{\widehat{\text{dCRT}}} \xrightarrow{p} 1$, so by conditional Slutsky's theorem (Theorem 3) we conclude that

$$
(\text{S32}) \quad \frac{1}{\widehat{S}_n^{\widehat{\text{dCRT}}}} T_n^{\widehat{\text{dCRT}}}(\widetilde{X}, X, Y, Z) = \frac{S_n^{\widehat{\text{dCRT}}}}{\widehat{S}_n^{\widehat{\text{dCRT}}}}\left(\frac{I_n}{S_n^{\widehat{\text{dCRT}}}} - \frac{J_n}{S_n^{\widehat{\text{dCRT}}}}\right)\bigg|\, X, Y, Z \xrightarrow{d,p} N(0,1),
$$

as desired. $\qquad\square$

COROLLARY 3. *Under the assumptions of Theorem 6, if the non-accumulation condition* (15) *is satisfied then the* $\widehat{\text{dCRT}}$ *is asymptotically equivalent to the* $\widehat{\text{MX(2)}}$ *F-test.*

PROOF. The desired equivalence is given by Lemma 3 with $T_n(X, Y, Z) \equiv (\widehat{S}_n^{\widehat{\text{dCRT}}})^{-1} T_n^{\widehat{\text{dCRT}}}$ and $C_n(X, Y, Z) \equiv C_n^{\widehat{\text{dCRT}}}(X, Y, Z)$, since ϕ_n^1 and ϕ_n^2 in the lemma statement reduce to $\widehat{\text{dCRT}}$ and the $\widehat{\text{MX(2)}}$ F-test, respectively. This lemma is applicable because the convergence of the critical value (S22) is given by Theorem 6 and the non-accumulation condition (S23) is assumed. $\qquad\square$

APPENDIX D: PROOFS FOR SECTION 4 AND APPENDIX A

For the sake of this section, we define

$$
(\text{S33}) \quad s_n^2 \equiv \mathbb{E}_{\mathcal{L}_n}[\text{Var}_{\mathcal{L}_n}[\boldsymbol{X}|\boldsymbol{Z}]\text{Var}_{\mathcal{L}_n}[\boldsymbol{Y}|\boldsymbol{Z}]].
$$

D.1. Auxiliary lemmas.

LEMMA 4 (Conditional Jensen inequality, Davidson (2003), Theorem 10.18). *Let W be a random variable and let ϕ be a convex function, such that W and $\phi(W)$ are integrable. For any σ-algebra $\mathcal{F}$, we have the inequality*

$$
\phi(\mathbb{E}[W \mid \mathcal{F}]) \leq \mathbb{E}[\phi(W) \mid \mathcal{F}] \quad \text{almost surely.}
$$

LEMMA 5. *Let W_n be a sequence of random variables and $\mathcal{F}_n$ a sequence of σ-algebras. If $W_n \mid \mathcal{F}_n \xrightarrow{p,p} 0$, then $W_n \xrightarrow{p} 0$.*

PROOF. Let $\epsilon > 0$. Because of the assumed conditional convergence in probability, we have

$$
\mathbb{P}[|W_n| > \epsilon \mid \mathcal{F}_n] \xrightarrow{p} 0.
$$

By the bounded convergence theorem, it follows that

$$\mathbb{P}[|W_n| > \epsilon] = \mathbb{E}[\mathbb{P}[|W_n| > \epsilon \mid \mathcal{F}_n]] \to 0,$$

from which the conclusion follows. □

LEMMA 6. *Consider a sequence of laws $\mathcal{L}_n \in \mathscr{L}_n^0$. Given assumption* (SP2), *we have*

(S34)
$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}_{\mathcal{L}_n}[X_i \mid Z_i]\mathrm{Var}_{\mathcal{L}_n}[Y_i \mid Z_i] - s_n^2 \overset{p}{\to} 0;$$

(S35)
$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_{n,x}(Z_i))^2\mathrm{Var}_{\mathcal{L}_n}[Y_i \mid Z_i] - s_n^2 \overset{p}{\to} 0;$$

(S36)
$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}_{\mathcal{L}_n}[X_i \mid Z_i](Y_i - \mu_{n,y}(Z_i))^2 - s_n^2 \overset{p}{\to} 0.$$

Given additionally assumption (SP1), *we also have*

(S37)
$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \widehat{\mu}_{n,x}(Z_i))^2\mathrm{Var}_{\mathcal{L}_n}[Y_i \mid Z_i] - s_n^2 \overset{p}{\to} 0;$$

(S38)
$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}_{\mathcal{L}_n}[X_i \mid Z_i](Y_i - \widehat{\mu}_{n,y}(Z_i))^2 - s_n^2 \overset{p}{\to} 0.$$

Given additionally the conditional Lyapunov condition (Lyap-2) *and the variance consistency condition* (21), *we also have*

(S39)
$$(S_n^{\widehat{\mathrm{dCRT}}})^2 - s_n^2 \equiv \frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}_{\widehat{\mathcal{L}}_n}[X_i \mid Z_i](Y_i - \mu_{n,y}(Z_i))^2 - s_n^2 \overset{p}{\to} 0.$$

Given additionally the assumption $\widehat{E}'_{n,y} \overset{p}{\to} 0$, we also have

(S40)
$$(\widehat{S}_n^{\widehat{\mathrm{dCRT}}})^2 - s_n^2 \equiv \frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}_{\widehat{\mathcal{L}}_n}[X_i \mid Z_i](Y_i - \widehat{\mu}_{n,y}(Z_i))^2 - s_n^2 \overset{p}{\to} 0.$$

PROOF. We prove the convergence statements in order.

Proofs of statements (S34), (S35), (S36). These statements are consequences of the weak law of large numbers (Corollary 2). To verify the statement (S34), we note that

$$\mathbb{E}_{\mathcal{L}_n}[\mathrm{Var}_{\mathcal{L}_n}[X_i \mid Z_i]\mathrm{Var}_{\mathcal{L}_n}[Y_i \mid Z_i]] = \mathbb{E}_{\mathcal{L}_n}[\mathrm{Var}_{\mathcal{L}_n}[\boldsymbol{X} \mid \boldsymbol{Z}]\mathrm{Var}_{\mathcal{L}_n}[\boldsymbol{Y} \mid \boldsymbol{Z}]] \equiv s_n^2,$$

and

$$\sup_{i,n} \mathbb{E}_{\mathcal{L}_n}[|\mathrm{Var}_{\mathcal{L}_n}[X_i \mid Z_i]\mathrm{Var}_{\mathcal{L}_n}[Y_i \mid Z_i]|^{1+\delta/2}]$$

$$= \sup_{n} \mathbb{E}_{\mathcal{L}_n}[|\mathrm{Var}_{\mathcal{L}_n}[\boldsymbol{X} \mid \boldsymbol{Z}]\mathrm{Var}_{\mathcal{L}_n}[\boldsymbol{Y} \mid \boldsymbol{Z}]|^{1+\delta/2}]$$

(S41)
$$\leq \sup_{n} \mathbb{E}_{\mathcal{L}_n}[\mathbb{E}_{\mathcal{L}_n}[|\boldsymbol{X} - \mu_{n,x}(\boldsymbol{Z})|^{2+\delta} \mid \boldsymbol{Z}]\mathbb{E}_{\mathcal{L}_n}[|\boldsymbol{Y} - \mu_{n,y}(\boldsymbol{Z})|^{2+\delta} \mid \boldsymbol{Z}]]$$

$$= \sup_{n} \mathbb{E}_{\mathcal{L}_n}[|(\boldsymbol{X} - \mu_{n,x}(\boldsymbol{Z}))(\boldsymbol{Y} - \mu_{n,y}(\boldsymbol{Z}))|^{2+\delta}]$$

$$< c_2 < \infty.$$

The inequality in the third line follows from the conditional Jensen inequality (Lemma 4), the equality in the fourth line follows from the conditional independence assumption, and the inequality in the fifth line follows from the $2 + \delta$ moment assumption (SP2). Hence we have verified the sufficient condition (S13) for the WLLN, so the convergence (S34) follows. Statements (S35) and (S36) can be verified with similar arguments.

Proofs of statements (S37) *and* (S38). We prove only the first, as the second will follow by symmetry. Given the convergence (S35), the statement (S37) will follow if we show that

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \widehat{\mu}_{n,x}(Z_i))^2\mathrm{Var}_{\mathcal{L}_n}[Y_i \mid Z_i] - \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_{n,x}(Z_i))^2\mathrm{Var}_{\mathcal{L}_n}[Y_i \mid Z_i]$$

$$= \frac{1}{n}\sum_{i=1}^{n}(\widehat{\mu}_{n,x}(Z_i) - \mu_{n,x}(Z_i))^2\mathrm{Var}_{\mathcal{L}_n}[Y_i \mid Z_i]$$

(S42)
$$- \frac{2}{n}\sum_{i=1}^{n}(X_i - \mu_{n,x}(Z_i))(\widehat{\mu}_{n,x}(Z_i) - \mu_{n,x}(Z_i))\mathrm{Var}_{\mathcal{L}_n}[Y_i \mid Z_i]$$

$$\equiv (E'_{n,x})^2 - \frac{2}{n}\sum_{i=1}^{n}(X_i - \mu_{n,x}(Z_i))(\widehat{\mu}_{n,x}(Z_i) - \mu_{n,x}(Z_i))\mathrm{Var}_{\mathcal{L}_n}[Y_i \mid Z_i]$$

$$\xrightarrow{p} 0.$$

The convergence $E'_{n,x} \xrightarrow{p} 0$ is assumed (SP1). To verify the convergence of the second term, note first that conditional Hölder (Lemma 12) and the derivation (S41) give

(S43)
$$\sup_n s_n^2 \leq \sup_n \{\mathbb{E}_{\mathcal{L}_n}[|\mathrm{Var}_{\mathcal{L}_n}[\boldsymbol{X} \mid \boldsymbol{Z}]\mathrm{Var}_{\mathcal{L}_n}[\boldsymbol{Y} \mid \boldsymbol{Z}]|^{1+\delta/2}]\}^{2/(2+\delta)} < \infty,$$

which combined with the convergence (S35) implies that

(S44)
$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_{n,x}(Z_i))^2\mathrm{Var}_{\mathcal{L}_n}[Y_i \mid Z_i] = O_p(1).$$

Therefore, by the Cauchy-Schwartz inequality we find that

$$\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_{n,x}(Z_i))(\widehat{\mu}_{n,x}(Z_i) - \mu_{n,x}(Z_i))\mathrm{Var}_{\mathcal{L}_n}[Y_i \mid Z_i]\right)^2$$

$$\leq \left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_{n,x}(Z_i))^2\mathrm{Var}_{\mathcal{L}_n}[Y_i \mid Z_i]\right)\left(\frac{1}{n}\sum_{i=1}^{n}(\widehat{\mu}_{n,x}(Z_i) - \mu_{n,x}(Z_i))^2\mathrm{Var}_{\mathcal{L}_n}[Y_i \mid Z_i]\right)$$

$$\equiv \left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_{n,x}(Z_i))^2\mathrm{Var}_{\mathcal{L}_n}[Y_i \mid Z_i]\right) \cdot (E'_{n,x})^2$$

$$= O_p(1) \cdot o_p(1) = o_p(1).$$

This proves the convergence (S42), which in turn implies the claimed convergence (S37).

Proof of statement (S39). Note that

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}_{\widehat{\mathcal{L}}_n}[X_i \mid Z_i](Y_i - \mu_{n,y}(Z_i))^2 - s_n^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}_{\widehat{\mathcal{L}}_n}[X_i \mid Z_i](Y_i - \mu_{n,y}(Z_i))^2 - \frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}_{\widehat{\mathcal{L}}_n}[X_i \mid Z_i]\mathrm{Var}[Y_i \mid Z_i]$$

$$+ \frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}_{\widehat{\mathcal{L}}_n}[X_i \mid Z_i]\mathrm{Var}[Y_i \mid Z_i] - \frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}_{\mathcal{L}_n}[X_i \mid Z_i]\mathrm{Var}[Y_i \mid Z_i]$$

$$+ \frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}_{\mathcal{L}_n}[X_i \mid Z_i]\mathrm{Var}[Y_i \mid Z_i] - s_n^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}_{\widehat{\mathcal{L}}_n}[X_i \mid Z_i](Y_i - \mu_{n,y}(Z_i))^2 - \frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}_{\widehat{\mathcal{L}}_n}[X_i \mid Z_i]\mathrm{Var}[Y_i \mid Z_i] + o_p(1),$$

where we used the variance consistency assumption (21) and the convergence result (S34) to obtain the last line. Hence, it suffices to show that

$$\text{(S45)} \qquad \frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{Var}_{\widehat{\mathcal{L}}_n}[X_i \mid Z_i](Y_i - \mu_{n,y}(Z_i))^2 - \mathrm{Var}_{\widehat{\mathcal{L}}_n}[X_i \mid Z_i]\mathrm{Var}[Y_i \mid Z_i]\right) \xrightarrow{p} 0.$$

To this end, we apply the conditional WLLN (Theorem 4) with $\mathcal{F}_n = \sigma(X, Z)$ and

$$W_{in} = \mathrm{Var}_{\widehat{\mathcal{L}}_n}[X_i \mid Z_i](Y_i - \mu_{n,y}(Z_i))^2.$$

We check the required $1 + \delta$ moment condition (S8):

$$\frac{1}{n^{1+\delta/2}}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{L}_n}[|W_{in}|^{1+\delta/2} \mid \mathcal{F}_n]$$

$$\equiv \frac{1}{n^{1+\delta/2}}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{L}_n}\left[\mathrm{Var}_{\widehat{\mathcal{L}}_n}[X_i \mid Z_i]^{1+\delta/2} \cdot |Y_i - \mu_{n,y}(Z_i)|^{2+\delta} \mid X, Z\right]$$

$$\leq \frac{1}{n^{1+\delta/2}}\sum_{i=1}^{n}\mathbb{E}_{\mathcal{L}_n}\left[\mathbb{E}_{\widehat{\mathcal{L}}_n}[|\widetilde{X}_i - \widehat{\mu}_{n,x}(Z_i)|^{2+\delta} \mid X, Z] \cdot |Y_i - \mu_{n,y}(Z_i)|^{2+\delta} \mid X, Z\right]$$

$$= \frac{1}{n^{1+\delta/2}}\sum_{i=1}^{n}\mathbb{E}_{\widehat{\mathcal{L}}_n}[|\widetilde{X}_i - \widehat{\mu}_{n,x}(Z_i)|^{2+\delta} \mid X, Z] \cdot \mathbb{E}_{\mathcal{L}_n}\left[|Y_i - \mu_{n,y}(Z_i)|^{2+\delta} \mid Z_i\right]$$

$$\xrightarrow{p} 0.$$

The inequality in the third line follows from the conditional Jensen inequality (Lemma 4), the equality in the fourth line from the assumed conditional independence, and the convergence in the fifth line from the conditional Lyaponov assumption (Lyap-2). Therefore, the conditional WLLN gives

$$\frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{Var}_{\widehat{\mathcal{L}}_n}[X_i \mid Z_i](Y_i - \mu_{n,y}(Z_i))^2 - \mathrm{Var}_{\widehat{\mathcal{L}}_n}[X_i \mid Z_i]\mathrm{Var}_{\mathcal{L}_n}[Y_i \mid Z_i]\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{Var}_{\widehat{\mathcal{L}}_n}[X_i \mid Z_i](Y_i - \mu_{n,y}(Z_i))^2 - \mathbb{E}_{\mathcal{L}_n}[\mathrm{Var}_{\widehat{\mathcal{L}}_n}[X_i \mid Z_i](Y_i - \mu_{n,y}(Z_i))^2 \mid X, Z]\right)$$

$$\xrightarrow{p,p} 0,$$

where the equality follows from the assumed conditional independence. Since conditional convergence in probability implies unconditional convergence in probability (Lemma 5), this verifies the claimed convergence statement (S45) and completes the proof of the statement (S39).

Proof of statement (S40). Given the convergence (S39), it suffices to show that

$$\text{(S46)} \quad \frac{1}{n}\sum_{i=1}^{n}\text{Var}_{\widehat{\mathcal{L}}_n}[X_i \mid Z_i](Y_i - \widehat{\mu}_{n,y}(Z_i))^2 - \frac{1}{n}\sum_{i=1}^{n}\text{Var}_{\widehat{\mathcal{L}}_n}[X_i \mid Z_i](Y_i - \mu_{n,y}(Z_i))^2 \xrightarrow{p} 0$$

Given the assumption that $\widehat{E}'_{n,y} \xrightarrow{p} 0$, this statement's proof is analogous to that of statement (S42), so we omit it for the sake of brevity. This completes the proof of the lemma. □

LEMMA 7. *Define*

$$\text{(S47)} \quad (S_n^{\widehat{\text{dCRT}}})^2 \equiv \frac{1}{n}\sum_{i=1}^{n}\text{Var}_{\widehat{\mathcal{L}}_n}[X_i|Z_i](Y_i - \mu_{n,y}(Z_i))^2.$$

Under the assumptions of Theorem 2, we have

$$\text{(S48)} \quad \frac{(S_n^{\widehat{\text{dCRT}}})^2}{(\widehat{S}_n^{\widehat{\text{dCRT}}})^2} \xrightarrow{p} 1$$

and

$$\text{(S49)} \quad \mathbb{P}[(S_n^{\widehat{\text{dCRT}}})^2 > \epsilon] \to 1 \quad \textit{for some } \epsilon > 0.$$

PROOF. The equivalence of variances (S48) follows from the convergences (S39) and (S40) (Lemma 6), as well as the observation that conditional independence and the assumption (SP2) imply

$$\text{(S50)} \quad \inf_n s_n^2 = \inf_n \mathbb{E}_{\mathcal{L}_n}[(\boldsymbol{X} - \mu_{n,x}(\boldsymbol{Z}))^2(\boldsymbol{Y} - \mu_{n,y}(\boldsymbol{Z}))^2] > 0.$$

The stochastic boundedness from below (S49) follows from the latter fact and the convergence (S39). □

LEMMA 8. *If the nondegeneracy conditions* (NDG1) *and* (NDG2) *and the conditional Lyapunov condition* (Lyap-1) *hold, then the variance estimates* $(\widehat{S}_n^{\text{GCM}})^2$ *and* $(\widehat{S}_n^{\widehat{\text{dCRT}}})^2$ *are equivalent under resampling:*

$$\text{(S51)} \quad \frac{(\widehat{S}_n^{\widehat{\text{dCRT}}}(X,Y,Z))^2}{(\widehat{S}_n^{\text{GCM}}(\widetilde{X},X,Y,Z))^2} \xrightarrow{p} 1.$$

PROOF. Define

$$W_{in} \equiv (\widetilde{X}_i - \widehat{\mu}_{n,x}(Z_i))(Y_i - \widehat{\mu}_{n,y}(Z_i)),$$

so that

$$(\widehat{S}_n^{\text{GCM}}(\widetilde{X},X,Y,Z))^2 \equiv \frac{1}{n}\sum_{i=1}^{n}W_{in}^2 - \left(\frac{1}{n}\sum_{i=1}^{n}W_{in}\right)^2.$$

First we claim that $\frac{1}{n}\sum_{i=1}^{n} W_{in} \overset{p}{\to} 0$. We will use conditional WLLN (Theorem 4) with $\mathcal{F}_n \equiv \sigma(X,Y,Z)$. First note that $\mathbb{E}[W_{in} \mid \mathcal{F}_n] = 0$ by construction. We also check the moment condition (S8):

$$\frac{1}{n^{2+\delta}}\sum_{i=1}^{n}\mathbb{E}[|W_{in}|^{2+\delta} \mid \mathcal{F}_n]$$

$$= \frac{1}{n^{1+\delta/2}}\frac{1}{n^{1+\delta/2}}\sum_{i=1}^{n}|Y_i - \widehat{\mu}_{n,y}(Z_i)|^{2+\delta}\mathbb{E}\left[|\widetilde{X}_i - \widehat{\mu}_{n,x}(Z_i)|^{2+\delta} \mid X, Z\right] \overset{p}{\to} 0,$$

where the latter convergence is by the assumption (Lyap-1). Hence we have that $\frac{1}{n}\sum_{i=1}^{n} W_{in} \mid \mathcal{F}_n \overset{p,p}{\to} 0$ which by Lemma 5 implies $\frac{1}{n}\sum_{i=1}^{n} W_{in} \overset{p}{\to} 0$.

Next we show that $\frac{1}{n}\sum_{i=1}^{n} W_{in}^2 - (\widehat{S}_n^{\widehat{\text{dCRT}}})^2 \overset{p}{\to} 0$. We will use conditional WLLN with $\mathcal{F}_n = \sigma(X,Y,Z)$, observing that $\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} W_{in}^2 \mid \mathcal{F}_n\right] = (\widehat{S}_n^{\widehat{\text{dCRT}}})^2$. Next we verify the moment condition (S8):

$$(\text{S52}) \quad \frac{1}{n^{1+\delta/2}}\sum_{i=1}^{n}\mathbb{E}[|W_{in}|^{2+\delta} \mid \mathcal{F}_n]$$

$$(\text{S53}) \qquad = \frac{1}{n^{1+\delta/2}}\sum_{i=1}^{n}|Y_i - \widehat{\mu}_{n,y}(Z_i)|^{2+\delta}\mathbb{E}\left[|\widetilde{X}_i - \widehat{\mu}_{n,x}(Z_i)|^{2+\delta} \mid X, Z\right] \overset{p}{\to} 0,$$

where the latter convergence is by the assumption (Lyap-1). Hence we have that $\frac{1}{n}\sum_{i=1}^{n} W_{in}^2 - (\widehat{S}_n^{\widehat{\text{dCRT}}})^2 \mid \mathcal{F}_n \overset{p,p}{\to} 0$ which by Lemma 5 implies that $\frac{1}{n}\sum_{i=1}^{n} W_{in}^2 - (\widehat{S}_n^{\widehat{\text{dCRT}}})^2 \overset{p}{\to} 0$.

Combining both of these results we find that

$$(\widehat{S}_n^{\text{GCM}}(\widetilde{X}, X, Y, Z))^2 - (\widehat{S}_n^{\widehat{\text{dCRT}}}(X, Y, Z))^2 \overset{p}{\to} 0.$$

Now using the nondegeneracy condition (NDG1) we can conclude that (S51) holds true, as desired. □

D.2. Proofs of main results.

PROOF OF THEOREM 2. To show the asymptotic equivalence of variance estimates (24) it suffices to show that

$$(\text{S54}) \qquad (\widehat{S}_n^{\widehat{\text{dCRT}}})^2 - s_n^2 \overset{p}{\to} 0; \quad (\widehat{S}_n^{\text{GCM}})^2 - s_n^2 \overset{p}{\to} 0; \quad \inf_n s_n^2 > 0.$$

The first of these statements is given by equation (S40) in Lemma 6, the second follows from the proof of Theorem 6 in Shah and Peters (2020), and the third is a consequence of assumption (SP2) and conditional independence.

Given the asymptotic equivalence of the variance estimates (24), we can show using Lemma 3 that the GCM test is asymptotically equivalent to the $\widehat{\text{MX(2)}}$ F-test. Indeed, set

$$(\text{S55}) \qquad T_n \equiv T_n^{\text{GCM}} \quad \text{and} \quad C_n \equiv \frac{\widehat{S}_n^{\widehat{\text{dCRT}}}}{\widehat{S}_n^{\text{GCM}}}z_{1-\alpha}$$

Then, ϕ_n^1 is the $\widehat{\text{MX(2)}}$ F-test and ϕ_n^2 is the GCM test. The asymptotic equivalence of the variance estimates (24) then implies the critical value convergence assumption (S22) of Lemma 3. The non-accumulation assumption (S23) is a consequence of the fact that, under the assumptions of Theorem 2, we have $T_n^{\text{GCM}} \overset{d}{\to} N(0,1)$ (Shah and Peters, 2020). On the other hand,

by Corollary 3 (whose conclusion holds under the assumptions of Theorem 2), we also know that $\widehat{\mathrm{dCRT}}$ is asymptotically equivalent to the $\widehat{\mathrm{MX(2)}}$ F-test. Hence, both the GCM test and the $\widehat{\mathrm{dCRT}}$ are asymptotically equivalent to the $\widehat{\mathrm{MX(2)}}$ F-test, so they are asymptotically equivalent to each other as well. □

PROOF OF PROPOSITION 1. We treat the two cases separately.

Case 1. We have

$$\frac{1}{n}\sum_{i=1}^{n}((X_i - \widehat{\mu}_{n,x}(Z_i))^2 - \mathrm{Var}_{\mathcal{L}_n}[X_i \mid Z_i])\mathrm{Var}_{\mathcal{L}_n}[Y_i \mid Z_i]$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i - \widehat{\mu}_{n,x}(Z_i))^2\mathrm{Var}_{\mathcal{L}_n}[Y_i \mid Z_i] - \frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}_{\mathcal{L}_n}[X_i \mid Z_i]\mathrm{Var}_{\mathcal{L}_n}[Y_i \mid Z_i]$$

$$= (s_n^2 + o_p(1)) - (s_n^2 + o_p(1))$$

$$= o_p(1),$$

where the third line follows from the convergences (S37) and (S34) from Lemma 6. This shows the variance consistency property (21).

Case 2. We need to show that

$$\frac{1}{n}\sum_{i=1}^{n}(f(\widehat{\mu}_{n,x}(Z_i)) - f(\mu_{n,x}(X_i)))\mathrm{Var}_{\mathcal{L}_n}[Y_i \mid Z_i] \xrightarrow{p} 0.$$

By the Cauchy-Schwartz inequality, we have

(S56)
$$\left|\frac{1}{n}\sum_{i=1}^{n}(f(\widehat{\mu}_{n,x}(Z_i)) - f(\mu_{n,x}(Z_i)))\mathrm{Var}_{\mathcal{L}_n}[Y_i \mid Z_i]\right|$$

$$\leq \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f(\widehat{\mu}_{n,x}(Z_i)) - f(\mu_{n,x}(Z_i))^2\mathrm{Var}_{\mathcal{L}_n}[Y_i \mid Z_i]}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}_{\mathcal{L}_n}[Y_i \mid Z_i]}.$$

Given the assumption that $\sup_n \mathbb{E}_{\mathcal{L}_n}[|\boldsymbol{Y} - \mu_{n,y}(\boldsymbol{Z})|^{2+\delta}] < \infty$ for some $\delta > 0$, Jensen's inequality gives

$$\sup_n \mathbb{E}[\mathrm{Var}_{\mathcal{L}_n}[\boldsymbol{Y} \mid \boldsymbol{Z}]^{1+\delta/2}] \leq \sup_n \mathbb{E}_{\mathcal{L}_n}[|\boldsymbol{Y} - \mu_{n,y}(\boldsymbol{Z})|^{2+\delta}] < \infty.$$

Therefore, the weak law of large numbers (Corollary 2) gives

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}_{\mathcal{L}_n}[Y_i \mid Z_i] - \mathbb{E}[\mathrm{Var}_{\mathcal{L}_n}[\boldsymbol{Y} \mid \boldsymbol{Z}]] \xrightarrow{p} 0.$$

Furthermore,

$$\mathbb{E}[\mathrm{Var}_{\mathcal{L}_n}[\boldsymbol{Y} \mid \boldsymbol{Z}]] \leq \sup_n \mathbb{E}_{\mathcal{L}_n}[|\boldsymbol{Y} - \mu_{n,y}(\boldsymbol{Z})|^{2+\delta}]^{\frac{2}{2+\delta}} < \infty,$$

so

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}_{\mathcal{L}_n}[Y_i \mid Z_i] = O_p(1).$$

On the other hand, we know $\text{supp}(\mu_{n,x}(Z_i)) \subseteq \text{Conv}(\text{supp}(\mathcal{L}_n(\boldsymbol{X}))$ for every i and n and by assumption $\widehat{\mu}_{n,x}(\boldsymbol{Z}) \subseteq \text{Conv}(\text{supp}(\mathcal{L}_n(\boldsymbol{X})))$ almost surely. Together with the fact that f is Lipschitz (say with Lipschitz constant L) on $\cup_{n=1}^\infty \text{Conv}(\text{supp}(\mathcal{L}_n(\boldsymbol{X})))$, it follows that

$$\frac{1}{n} \sum_{i=1}^n (f(\widehat{\mu}_{n,x}(Z_i)) - f(\mu_{n,x}(Z_i)))^2 \text{Var}_{\mathcal{L}_n}[Y_i \mid Z_i]$$

$$\leq \frac{L^2}{n} \sum_{i=1}^n (\widehat{\mu}_{n,x}(Z_i) - \mu_{n,x}(Z_i))^2 \text{Var}_{\mathcal{L}_n}[Y_i \mid Z_i] \equiv L^2 (E'_{n,x})^2 = o_p(1).$$

Combining the last two displays with equation (S56) gives us the desired result. $\qquad\square$

PROOF OF COROLLARY 2. Define the event

$$A_n \equiv \{\phi_n^{\widehat{\text{dCRT}}}(X, Y, Z) \neq \phi_n^{\text{GCM}}(X, Y, Z)\}$$

The conclusion (25) of Theorem 2 implies that $\mathbb{P}_{\mathcal{L}_n}[A_n] \to 0$, so by the assumed contiguity of $\mathcal{L}'_n$ to $\mathcal{L}_n$ we also have $\mathbb{P}_{\mathcal{L}'_n}[A_n] \to 0$. This shows the desired asymptotic equivalence of tests (26). To show the asymptotic equivalence of powers, we derive

$$|\mathbb{E}_{\mathcal{L}'_n}[\phi_n^{\widehat{\text{dCRT}}}(X, Y, Z) - \phi_n^{\text{GCM}}(X, Y, Z)]|$$

$$\leq \mathbb{E}_{\mathcal{L}'_n}[|\phi_n^{\widehat{\text{dCRT}}}(X, Y, Z) - \phi_n^{\text{GCM}}(X, Y, Z)|]$$

$$= \mathbb{E}_{\mathcal{L}'_n}[|\phi_n^{\widehat{\text{dCRT}}}(X, Y, Z) - \phi_n^{\text{GCM}}(X, Y, Z)|\mathbb{1}(\phi_n^{\widehat{\text{dCRT}}} \neq \phi_n^{\text{GCM}})]$$

$$\leq \mathbb{E}_{\mathcal{L}'_n}[\mathbb{1}(\phi_n^{\widehat{\text{dCRT}}} \neq \phi_n^{\text{GCM}})]$$

$$= \mathbb{P}_{\mathcal{L}'_n}[\phi_n^{\widehat{\text{dCRT}}} \neq \phi_n^{\text{GCM}}] \to 0.$$

This completes the proof. $\qquad\square$

PROOF OF COROLLARY 3. Fix $\epsilon > 0$, and for each n let $\mathcal{L}_n^* \in \mathscr{L}_n^0 \cap \mathscr{R}_n$ be such that

$$\mathbb{P}_{\mathcal{L}_n^*}[\phi_n^{\widehat{\text{dCRT}}}(X, Y, Z) \neq \phi_n^{\text{GCM}}(X, Y, Z)]$$

$$\geq \sup_{\mathcal{L}_n \in \mathscr{L}_n^0 \cap \mathscr{R}_n} \mathbb{P}_{\mathcal{L}_n}[\phi_n^{\widehat{\text{dCRT}}}(X, Y, Z) \neq \phi_n^{\text{GCM}}(X, Y, Z)] - \epsilon.$$

Applying Theorem 2 to the sequence $\mathcal{L}_n^*$ and using the asymptotic Type-I error control of the GCM test, we obtain

$$\limsup_{n \to \infty} \sup_{\mathcal{L}_n \in \mathscr{L}_n^0 \cap \mathscr{R}_n} \mathbb{E}_{\mathcal{L}_n}[\phi_n^{\widehat{\text{dCRT}}}(X, Y, Z)]$$

$$\leq \limsup_{n \to \infty} \sup_{\mathcal{L}_n \in \mathscr{L}_n^0 \cap \mathscr{R}_n} |\mathbb{E}_{\mathcal{L}_n}[\phi_n^{\widehat{\text{dCRT}}}(X, Y, Z)] - \mathbb{E}_{\mathcal{L}_n}[\phi_n^{\text{GCM}}(X, Y, Z)]| + \mathbb{E}_{\mathcal{L}_n}[\phi_n^{\text{GCM}}(X, Y, Z)]$$

$$\leq \limsup_{n \to \infty} \sup_{\mathcal{L}_n \in \mathscr{L}_n^0 \cap \mathscr{R}_n} |\mathbb{E}_{\mathcal{L}_n}[\phi_n^{\widehat{\text{dCRT}}}(X, Y, Z)] - \mathbb{E}_{\mathcal{L}_n}[\phi_n^{\text{GCM}}(X, Y, Z)]|$$

$$\quad + \limsup_{n \to \infty} \sup_{\mathcal{L}_n \in \mathscr{L}_n^0 \cap \mathscr{R}_n} \mathbb{E}_{\mathcal{L}_n}[\phi_n^{\text{GCM}}(X, Y, Z)]$$

$$\leq \limsup_{n \to \infty} \sup_{\mathcal{L}_n \in \mathscr{L}_n^0 \cap \mathscr{R}_n} \mathbb{P}_{\mathcal{L}_n}[\phi_n^{\widehat{\text{dCRT}}}(X, Y, Z) \neq \phi_n^{\text{GCM}}(X, Y, Z)]$$

$$\quad + \limsup_{n \to \infty} \sup_{\mathcal{L}_n \in \mathscr{L}_n^0 \cap \mathscr{R}_n} \mathbb{E}_{\mathcal{L}_n}[\phi_n^{\text{GCM}}(X, Y, Z)]$$

$$\leq \limsup_{n \to \infty} \sup_{\mathcal{L}_n \in \mathscr{L}_n^0 \cap \mathscr{R}_n} \mathbb{P}_{\mathcal{L}_n}[\phi_n^{\widehat{\text{dCRT}}}(X, Y, Z) \neq \phi_n^{\text{GCM}}(X, Y, Z)] + \alpha$$

$$\leq \limsup_{n \to \infty} \mathbb{P}_{\mathcal{L}_n^*}[\phi_n^{\widehat{\text{dCRT}}}(X, Y, Z) \neq \phi_n^{\text{GCM}}(X, Y, Z)] + \epsilon + \alpha$$

$$= \epsilon + \alpha.$$

Sending $\epsilon \to 0$ gives the desired conclusion. □

PROOF OF THEOREM 1. We have

$$T_n^{\widehat{\text{ndCRT}}}(\widetilde{X}, X, Y, Z) \equiv \frac{T_n^{\text{dCRT}}(\widetilde{X}, X, Y, Z)}{S_n^{\text{GCM}}(\widetilde{X}, X, Y, Z)} \equiv \frac{S_n^{\widehat{\text{dCRT}}}(X, Y, Z)}{S_n^{\text{GCM}}(\widetilde{X}, X, Y, Z)} \cdot \frac{T_n^{\text{dCRT}}(\widetilde{X}, X, Y, Z)}{S_n^{\widehat{\text{dCRT}}}(X, Y, Z)}$$

The first factor converges to 1 in probability (Lemma 8), whereas the second factor converges conditionally on X, Y, Z to $N(0, 1)$ (Theorem 1). Putting these two statements together with conditional Slutsky's theorem (Theorem 3), we arrive at the convergence (S2). Since the standard normal has continuous CDF we can use Lemma 1 to conclude the convergence of the critical value (S3). The equivalence statement (S4) follows from the convergence (S3) and Lemma 3 applied with $T_n = T_n^{\text{GCM}}$ and $C_n = C_n^{\widehat{\text{ndCRT}}}$. □

PROOF OF COROLLARY 1. The proof of this corollary is directly analogous to that of Corollary 3, so we omit it for the sake of brevity.

□

APPENDIX E: PROOFS FOR SECTION 5

The goal of this section is to prove our main optimality result (Theorem 3) and Corollary 4. The idea of the proof of Theorem 3 is to reduce the problem to a semiparametric testing problem, and then to use existing semiparametric optimality theory. To this end, we first review the relevant semiparametric theory (Section E.1). Then we leverage this theory to prove Theorem 3 (Section E.2) and verify Corollary 4 (Section E.3). Finally, we carry out deferred semiparametric computations (Section E.4).

E.1. Semiparametric preliminaries. Consider a semiparametric model parameterized by

(S57) $$(\beta, g) \in \mathbb{R} \times \mathcal{H}_g \subseteq \mathbb{R} \times L^2(\nu),$$

where ν is a measure on $\mathbb{R}^p$ and $\mathcal{H}_g \subseteq L^2(\nu)$ is a linear subspace. First, we define a notion of local Type-I error control within the context of the semiparametric model.

DEFINITION 2. Fix a point $g_0 \in \mathcal{H}_g$, and define $\theta_0 \equiv (0, g_0)$. A sequence of tests ϕ_n of $H_0 : \beta = 0$ has asymptotic Type-I error control at θ_0 relative to the tangent space $\dot{\mathcal{L}}_{\theta_0}$ if, for each submodel $t \mapsto \mathcal{L}_{(0,g_t)}$ with score in $\dot{\mathcal{L}}_{\theta_0}$ along which β is differentiable, we have

(S58) $$\limsup_{n \to \infty} \mathbb{E}_{\mathcal{L}_{(0, g_{1/\sqrt{n}})}}[\phi_n(W)] \leq \alpha.$$

This definition is most similar to that of Choi, Hall and Schick (1996), except the latter paper does not explicitly use the language of tangent spaces; our definition accommodates Type-I error control over more restricted sets of null distributions reflecting regularity conditions. Next we state a version of the classic semiparametric optimality result:

THEOREM 7 (Theorem 1 in Choi, Hall and Schick (1996), Theorem 25.44 in Van Der Vaart (1998), Theorem 18.12 in Kosorok (2008)). *Consider a semiparametric model $\{\mathcal{L}_{\beta,g} : (\beta, g) \in \mathbb{R} \times \mathcal{H}_g\}$ and a point $\theta_0 \equiv (0, g_0)$ for some $g_0 \in \mathcal{H}_g$. Suppose β is differentiable at $\mathcal{L}_{\theta_0}$ relative to the tangent space $\dot{\mathcal{L}}_{\theta_0}$ with efficient influence function $\widetilde{S}/\widetilde{I}(\theta_0)$, where $\widetilde{S}$ is the efficient score and $\widetilde{I}(\theta_0) > 0$ is the efficient information. For any sequence of tests ϕ_n of $H_0 : \beta = 0$ with asymptotic Type-I error control at θ_0 relative to the tangent space $\dot{\mathcal{L}}_{\theta_0}$ and any differentiable submodel $\mathcal{L}_t = \mathcal{L}_{(th_\beta, g_t)}$ with score in $\dot{\mathcal{L}}_{\theta_0}$ we have*

(S59) $$\limsup_{n \to \infty} \mathbb{E}_{\mathcal{L}_{1/\sqrt{n}}}[\phi_n(W)] \leq 1 - \Phi(z_{1-\alpha} - h_\beta \cdot \widetilde{I}(\theta_0)^{1/2}).$$

This bound is achieved by the efficient score test $\phi_n^{\mathrm{opt}}(X, Y, Z) \equiv \mathbb{1}(T_n^{\mathrm{opt}}(X, Y, Z) > z_{1-\alpha})$, where

(S60) $$T_n^{\mathrm{opt}}(X, Y, Z) \equiv \frac{1}{\widetilde{I}(\theta_0)^{1/2} n^{1/2}} \sum_{i=1}^{n} \widetilde{S}(X_i, Y_i, Z_i).$$

In other words,

(S61) $$\lim_{n \to \infty} \mathbb{E}_{\mathcal{L}_{1/\sqrt{n}}}[\phi_n^{\mathrm{opt}}(W)] = \lim_{n \to \infty} \mathbb{P}_{\mathcal{L}_{1/\sqrt{n}}}[T_n^{\mathrm{opt}}(X, Y, Z) > z_{1-\alpha}]$$
$$= 1 - \Phi(z_{1-\alpha} - h_\beta \cdot \widetilde{I}(\theta_0)^{1/2}).$$

This result is like Choi, Hall and Schick (1996, Theorem 1), except it explicitly deals with tangent spaces. On the other hand, the result is like Van Der Vaart (1998, Theorem 25.44) or Kosorok (2008, Theorem 18.12), except it is written in terms of semiparametric models and assumes Type-I error control in the sense of Definition 2 above. By comparison, Van Der Vaart (1998); Kosorok (2008) assume Type-I error control at each point $(0, g)$ for $g \in \mathcal{H}_g$. By inspection of the proof of Van Der Vaart (1998, Theorem 25.44), only local Type-I error control (Definition 2) is actually needed. In this sense, Theorem 7 can be verified using the same proof as that of Van Der Vaart (1998, Theorem 25.44), specializing to the case of semiparametric models.

E.2. Proof of Theorem 3. To apply the semiparametric theory from the previous section, the following lemma (proved in Section E.4) identifies the tangent space, the efficient score, and the efficient information at $\mathcal{L}_{\theta_0}$. These results are not novel or surprising; similar results are stated, for example, by Robins and Rotnitzky (2001) in the cases of linear, logistic, and Poisson regressions. Nevertheless, we state and prove Lemma 9 for a self-contained exposition and for precisely tracking the technical assumptions used.

LEMMA 9. [Proof] *In the context of the semiparametric model* (29)*, suppose the following assumptions hold:*

(S62) $$s^2(\theta_0) \equiv \mathbb{E}_{\mathcal{L}_{\theta_0}}[\mathrm{Var}_{\mathcal{L}_{\theta_0}}[\boldsymbol{X}|\boldsymbol{Z}]\mathrm{Var}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y}|\boldsymbol{Z}]] > 0;$$

(S63) $\quad \ddot{\psi} = K > 0$ *and* $\mathbb{E}_{\mathcal{L}_{x,z}}[\boldsymbol{X}^2] < \infty$ *OR* $\mathrm{supp}(\boldsymbol{X}, \boldsymbol{Z})$ *is compact and* $\mathcal{H}_g \subseteq C(\mathbb{R}^p),$

(S64) $$\mathbb{E}_{\mathcal{L}_{x,z}}[\boldsymbol{X}|\cdot] \in \mathcal{H}_g.$$

For each $h = (h_\beta, h_g) \in \mathbb{R} \times \mathcal{H}_g$, *the parametric submodel* $t \mapsto \mathcal{L}_{(th_\beta, g_0 + th_g)}$ *is differentiable in quadratic mean at* $t = 0$ *with score function*

(S65) $$S(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}) = \big(\boldsymbol{X}h_\beta + h_g(\boldsymbol{Z})\big)(\boldsymbol{Y} - \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y} \mid \boldsymbol{Z}])$$

and satisfies the following local asymptotic normality:

(S66) $$\log \prod_{i=1}^{n} \frac{\mathrm{d}\mathcal{L}_{\theta_n(h)}}{\mathrm{d}\mathcal{L}_{\theta_0}}(X_i, Y_i, Z_i) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} S(X_i, Y_i, Z_i) - \frac{1}{2}\mathrm{Var}_{\mathcal{L}_{\theta_0}}[S(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})] + o_{\mathcal{L}_{\theta_0}}(1).$$

The parameter β *is differentiable at* $\mathcal{L}_{\theta_0}$ *relative to the tangent space*

(S67) $$\dot{\mathcal{L}}_{\theta_0} \equiv \{\big(\boldsymbol{X}h_\beta + h_g(\boldsymbol{Z})\big)(\boldsymbol{Y} - \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y} \mid \boldsymbol{Z}]) : (h_\beta, h_g) \in \mathbb{R} \times \mathcal{H}_g\}$$

with efficient score function

(S68) $$\widetilde{S}(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}) = (\boldsymbol{X} - \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{X} \mid \boldsymbol{Z}])(\boldsymbol{Y} - \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y} \mid \boldsymbol{Z}]),$$

efficient information

(S69) $$\widetilde{I}(\theta_0) = s^2(\theta_0) \equiv \mathbb{E}_{\mathcal{L}_{\theta_0}}[\mathrm{Var}_{\mathcal{L}_{\theta_0}}[\boldsymbol{X}|\boldsymbol{Z}]\mathrm{Var}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y}|\boldsymbol{Z}]],$$

and efficient influence function equal to the ratio of the efficient information and the efficient score.

Note that assumptions (S63) and (S64) of Lemma 9 are the same as assumptions (36) and (37) of Theorem 3 in the main text; they are restated here for the reader's convenience. Using Lemma 9 in conjunction with Theorem 7, we can prove Theorem 3.

PROOF OF THEOREM 3. Let ϕ_n be a level α test of H_0 as defined in equation (32), and fix $g_0 \in \mathcal{S}$. By assumption (38), $\theta_n(0, h_g) \in \mathcal{R}$ for all $h_g \in \mathcal{H}_g$ for all sufficiently large n. Therefore, ϕ_n also has asymptotic Type-I error control at $\theta_0 \equiv (0, g_0)$ relative to the tangent space $\dot{\mathcal{L}}_{\theta_0}$ (S67) in the sense of Definition 2. Indeed, it suffices to take submodels $t \mapsto \mathcal{L}_{(0, g_t)}$ for $g_t = g_0 + th_g$ and $h_g \in \mathcal{H}_g$, so that $\mathcal{L}_{(0, g_{1/\sqrt{n}})} = \mathcal{L}_{\theta_n(0, h_g)}$. By Lemma 9 (applicable because its first assumption (S62) is implied by assumption (35) of Theorem 3 and its last two assumptions are also assumed by Theorem 3), the assumptions of Theorem 7 are met with efficient score $\widetilde{S}$ (S68) and efficient information $s^2(\theta_0)$ (S69), so taking submodels $t \mapsto \mathcal{L}_{(th_\beta, g_0 + th_g)}$ we find

(S70) $$\limsup_{n \to \infty} \mathbb{E}_{\mathcal{L}_{\theta_n(h)}}[\phi_n(X, Y, Z)] \leq 1 - \Phi(z_{1-\alpha} - h_\beta \cdot s(\theta_0)).$$

On the other hand, because $\mathcal{L}_{\theta_0} \in \mathscr{R}$, it follows that

$$T_n^{\text{GCM}}(X, Y, Z) = \frac{1}{s(\theta_0)\sqrt{n}} \sum_{i=1}^{n} (X_i - \mathbb{E}_{\mathcal{L}_{\theta_0}}[X_i \mid Z_i])(Y_i - \mathbb{E}_{\mathcal{L}_{\theta_0}}[Y_i \mid Z_i]) + o_{\mathcal{L}_{\theta_0}}(1)$$

(S71)
$$= \frac{1}{\widetilde{I}(\theta_0)^{1/2} n^{1/2}} \sum_{i=1}^{n} \widetilde{S}(X_i, Y_i, Z_i) + o_{\mathcal{L}_{\theta_0}}(1)$$

$$= T_n^{\text{opt}}(X, Y, Z) + o_{\mathcal{L}_{\theta_0}}(1).$$

The first equality follows from the proof of Theorem 6 in Shah and Peters (2020), the second follows from the derivations of the efficient score (S68) and efficient information (S69) in Lemma 9, and the third from equation (S60) in Theorem 7. From the local asymptotic normality (S66) it follows that $\prod_{i=1}^{n} \mathcal{L}_{\theta_n(h)}$ and $\prod_{i=1}^{n} \mathcal{L}_{\theta_0}$ are contiguous by Le Cam's first lemma (Van Der Vaart, 1998, Example 6.5). It follows that

$$T_n^{\text{GCM}}(X, Y, Z) = T_n^{\text{opt}}(X, Y, Z) + o_{\mathcal{L}_{\theta_n(h)}}(1)$$

We therefore find that

$$1 - \Phi(z_{1-\alpha} - h_\beta \cdot s(\theta_0)) = \lim_{n\to\infty} \mathbb{P}_{\mathcal{L}_{\theta_n(h)}}[T_n^{\text{opt}} > z_{1-\alpha}] = \lim_{n\to\infty} \mathbb{P}_{\mathcal{L}_{\theta_n(h)}}[T_n^{\text{GCM}} > z_{1-\alpha}].$$

The first inequality follows from the conclusion (S61) of Theorem 7 and the second equality follows from equation (S71) and Le Cam's first lemma. Therefore we have shown that for any $h \in (0, \infty) \times \mathcal{H}_g$ and any level α conditional independence test ϕ_n, we have

$$\limsup_{n\to\infty} \mathbb{E}_{\mathcal{L}_{\theta_n(h)}}[\phi_n(X, Y, Z)] \leq 1 - \Phi(z_{1-\alpha} - h_\beta \cdot s(\theta_0)) = \lim_{n\to\infty} \mathbb{P}_{\mathcal{L}_{\theta_n(h)}}(T_n^{\text{GCM}} > z_{1-\alpha}).$$

This shows that ϕ_n^{GCM} is LAUMP(g_0) and verifies the claimed asymptotic power (39). Furthermore, since $g_0 \in \mathcal{S}$ was chosen arbitrarily, it follows that ϕ_n^{GCM} is also LAUMP($\mathcal{S}$). This completes the proof. $\square$

E.3. Proof of Corollary 4. It suffices to verify each of the four assumptions of Theorem 3.

Verification of assumption (35). Note that assumption (SP2) is satisfied because by construction, $\boldsymbol{X} - \mu_x(\boldsymbol{Z})$ and $\boldsymbol{Y} - \mu_y(\boldsymbol{Z})$ are independent standard normal random variables for any $\mathcal{L} \in \mathscr{R}$. Next, let

(S72)
$$k(z, z') = \sum_{j=1}^{\infty} \lambda_j e_j(z) e_j(z')$$

be an eigendecomposition of the Sobolev kernel k with eigenfunctions e_j orthonormal with respect to Unif$[0, 1]$. To verify assumption (SP1) given assumption (SP2), it suffices to prove the following statements (Shah and Peters, 2020, Theorem 11 and Remark 12):

$$\text{Var}_{\mathcal{L}}[\boldsymbol{X}|\boldsymbol{Z}], \text{Var}_{\mathcal{L}}[\boldsymbol{Y}|\boldsymbol{Z}] \leq \sigma^2 < \infty \quad \text{almost surely, for all } \mathcal{L} \in \mathscr{R};$$

$$\sup_{\mathcal{L} \in \mathscr{R}} \max(\|\mu_{n,x}\|_{W^{1,2}[0,1]}, \|\mu_{n,y}\|_{W^{1,2}[0,1]}) < \infty;$$

$$\sum_{j=1}^{\infty} \lambda_j < \infty.$$

The first two of these statements follow directly from the construction of $\mathscr{R}$. The third follows from the eigendecomposition of the Sobolev kernel under the uniform measure on $[0, 1]$ (Wainwright, 2019, Example 12.23) with $\lambda_j = (\frac{2}{(2j-1)\pi})^2$.

Verification of assumption (36). Since we are using the normal exponential family, we have $\ddot{\psi} = 1$. Furthermore,

$$
\text{(S73)} \qquad \mathbb{E}_{\mathcal{L}_{x,z}}[\boldsymbol{X}^2] = \mathbb{E}_{\mathcal{L}_{x,z}}[\mathbb{E}_{\mathcal{L}_{x,z}}[\boldsymbol{X}^2|\boldsymbol{Z}]] = \mathbb{E}_{\mathcal{L}_{x,z}}[\mu_{0x}(\boldsymbol{Z})^2 + 1] < \infty,
$$

since $\mu_{0x} \in W^{(1,2)}([0,1]) \subseteq L^2([0,1])$.

Verification of assumption (37). By construction, $\mathbb{E}_{\mathcal{L}_{x,z}}[\boldsymbol{X}| \cdot] = \mu_{0x} \in W^{1,2}([0,1])$.

Verification of assumption (38). This assumption is a consequence of the openness of the ball $B_{W^{1,2}}(0,C)$ in $W^2([0,1])$.

E.4. Proof of Lemma 9.

Differentiability of parametric submodels. Consider the parametric submodel $t \mapsto \mathcal{L}_{(th_\beta, g_0 + th_g)}$ for some $(h_\beta, h_g) \in \mathbb{R} \times \mathcal{H}_g$, and denote

$$
\eta_t(x,z) \equiv xth_\beta + g_0(z) + th_g(z).
$$

Letting λ_y be the base measure of the exponential family f_η, we denote $\lambda \equiv \mathcal{L}_{x,z} \times \lambda_y$ and $\mathrm{d}\mathcal{L}_{(th_\beta, g_0 + th_g)}(x,y,z)/\mathrm{d}\lambda$ the density of the parametric model for $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ with respect to λ. According to Van Der Vaart (1998, Lemma 7.6), this submodel is differentiable in quadratic mean at $t = 0$ if the map

$$
t \mapsto \sqrt{\frac{\mathrm{d}\mathcal{L}_{(th_\beta, g_0 + th_g)}}{\mathrm{d}\lambda}(x,y,z)} = \sqrt{\frac{\mathrm{d}f_{\eta_t}}{\mathrm{d}\lambda_y}(y)} = \exp\left(\frac{y\eta_t - \psi(\eta_t)}{2}\right)
$$

is continuously differentiable at $t = 0$ for each $(x,y,z) \in \mathbb{R}^{1+1+p}$ and the elements of the Fisher information matrix are well-defined and continuous at $t = 0$. To show continuous differentiability of the square root density, we compute that

$$
\frac{\partial}{\partial t} \sqrt{\frac{\mathrm{d}\mathcal{L}_{(th_\beta, g_0 + th_g)}}{\mathrm{d}\lambda}(x,y,z)} = \exp\left(\frac{y\eta_t - \psi(\eta_t)}{2}\right) \cdot \frac{(y - \dot{\psi}(\eta_t))}{2} \cdot (xh_\beta + h_g(z)).
$$

The linearity of η_t in t and the smoothness of ψ imply the continuous differentiability of the above function in t.

Next consider the information matrix

$$
I_t \equiv \mathbb{E}_{\mathcal{L}_{(th_\beta, g_0 + th_g)}}\left[\left(\frac{\partial}{\partial t} \log \frac{\mathrm{d}\mathcal{L}_{(th_\beta, g_0 + th_g)}}{\mathrm{d}\lambda}(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})\right)^2\right]
$$

$$
= \mathbb{E}_{\mathcal{L}_{(th_\beta, g_0 + th_g)}}\left[\left(\frac{\partial}{\partial t}(\boldsymbol{Y}\eta_t(\boldsymbol{X}, \boldsymbol{Z}) - \psi(\eta_t(\boldsymbol{X}, \boldsymbol{Z})))\right)^2\right]
$$

$$
= \mathbb{E}_{\mathcal{L}_{(\beta_t, g_t)}}[(\boldsymbol{Y} - \dot{\psi}(\eta_t(\boldsymbol{X}, \boldsymbol{Z})))^2(\boldsymbol{X}h_\beta + h_g(\boldsymbol{Z}))^2]
$$

$$
= \mathbb{E}_{\mathcal{L}_{x,z}}\left[(\boldsymbol{X}h_\beta + h_g(\boldsymbol{Z}))^2\ddot{\psi}(\eta_t(\boldsymbol{X}, \boldsymbol{Z}))\right].
$$

We must show that I_t is well-defined and continuous at $t = 0$. By assumption (S63), either $\ddot{\psi} = K > 0$ and $\mathbb{E}_{\mathcal{L}_{x,z}}[\boldsymbol{X}^2] < \infty$ or $(\boldsymbol{X}, \boldsymbol{Z})$ is compactly supported and $\mathcal{H}_g \subseteq C(\mathbb{R}^p)$. If $\ddot{\psi} = K > 0$ and $\mathbb{E}_{\mathcal{L}_{x,z}}[\boldsymbol{X}^2] < \infty$, then we have

$$
I_0 = C\mathbb{E}_{\mathcal{L}_{x,z}}\left[(\boldsymbol{X}h_\beta + h_g(\boldsymbol{Z}))^2\right] \leq 2Kh_\beta^2\mathbb{E}_{\mathcal{L}_{x,z}}[\boldsymbol{X}^2] + 2K\mathbb{E}_{\mathcal{L}_{x,z}}[h_g^2(\boldsymbol{Z})]
$$

$$
= 2Kh_\beta^2\mathbb{E}_{\mathcal{L}_{x,z}}[\boldsymbol{X}^2] + 2K\|h_g\|_{L^2(\nu)}^2 < \infty.
$$

Note that, for the sake of this proof, we denote

(S74) $\qquad \nu \equiv \mathcal{L}_{x,z}(\boldsymbol{Z}), \quad$ so that $\quad \|h_g\|^2_{L^2(\nu)} \equiv \mathbb{E}_{\mathcal{L}_{x,z}}[h_g(\boldsymbol{Z})^2] < \infty \quad$ for $h_g \in \mathcal{H}_g.$

Hence, I_0 is well-defined. I_t is also continuous at $t = 0$ because it does not depend on t.

On the other hand, suppose $(\boldsymbol{X}, \boldsymbol{Z})$ is compactly supported and $\mathcal{H}_g \subseteq C(\mathbb{R}^p)$. The quantity inside the expectation defining I_0 is a bounded random variable, because the assumed continuity of h_g implies that this quantity is a continuous function of a random vector $(\boldsymbol{X}, \boldsymbol{Z})$ with compact support. Hence, I_0 is well-defined because it is the expectation of a bounded random variable. To show continuity of I_t, note that by the assumed continuity of h_g and compact support of $(\boldsymbol{X}, \boldsymbol{Z})$ we have $\sup_{t \leq 1} \eta_t(\boldsymbol{X}, \boldsymbol{Z}) \leq B < \infty$ almost surely. Therefore, for $t \leq 1$, we have

$$
\begin{aligned}
|I_t - I_0| &= \left| \mathbb{E}_{\mathcal{L}_{x,z}} \left[(\boldsymbol{X}h_\beta + h_g(\boldsymbol{Z}))^2 (\ddot{\psi}(\eta_t) - \ddot{\psi}(\eta_0)) \right] \right| \\
&\leq \mathbb{E}_{\mathcal{L}_{x,z}} \left[(\boldsymbol{X}h_\beta + h_g(\boldsymbol{Z}))^2 |\ddot{\psi}(\eta_t) - \ddot{\psi}(\eta_0)| \right] \\
&\leq \mathbb{E}_{\mathcal{L}_{x,z}} \left[(\boldsymbol{X}h_\beta + h_g(\boldsymbol{Z}))^2 \sup_{|b| \leq B} |\dddot{\psi}(b)| \cdot |\eta_t - \eta_0| \right] \\
&\leq \sup_{|b| \leq B} |\dddot{\psi}(b)| \cdot \mathbb{E}_{\mathcal{L}_{x,z}} \left[|\boldsymbol{X}h_\beta + h_g(\boldsymbol{Z})|^3 \right] \cdot t.
\end{aligned}
$$

We have $\sup_{|b| \leq B} |\dddot{\psi}(b)| < \infty$ because $\dddot{\psi}$ is a continuous function, and $\boldsymbol{X}h_\beta + h_g(\boldsymbol{Z})$ almost surely bounded as before. Therefore, we conclude that $|I_t - I_0| \to 0$ as $t \to 0$, so I_t is indeed continuous at 0.

Hence, we conclude by Van Der Vaart (1998, Lemma 7.6) that the parametric submodel $t \mapsto \mathcal{L}_{(th_\beta, g_0 + th_g)}$ is differentiable in quadratic mean at $t = 0$ with score function

$$
\begin{aligned}
S(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}) &= \frac{\partial}{\partial t} (\boldsymbol{Y}\eta_t(\boldsymbol{X}, \boldsymbol{Z}) - \psi(\eta_t(\boldsymbol{X}, \boldsymbol{Z}))) \Big|_{t=0} \\
&= (\boldsymbol{Y} - \dot{\psi}(\eta_t(\boldsymbol{X}, \boldsymbol{Z})))(\boldsymbol{X}h_\beta + h_g(\boldsymbol{Z})) \Big|_{t=0} \\
&= (\boldsymbol{X}h_\beta + h_g(\boldsymbol{Z}))(\boldsymbol{Y} - \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y}|\boldsymbol{Z}]),
\end{aligned}
$$

as claimed (S65).

Local asymptotic normality of parametric submodels. The local asymptotic normality of parametric submodels (S66) follows from the previously established quadratic mean differentiability (Van Der Vaart, 1998, Theorem 7.2).

Efficient score and information. For $(h_\beta, h_g) \in \mathbb{R} \times \mathcal{H}_g$, define the score operator

(S75)
$$
\begin{aligned}
A(h) &\equiv \left(\boldsymbol{X}h_\beta + h_g(\boldsymbol{Z}) \right)\left(\boldsymbol{Y} - \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y} \mid \boldsymbol{Z}] \right) \\
&= \boldsymbol{X}(\boldsymbol{Y} - \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y} \mid \boldsymbol{Z}])h_\beta + (\boldsymbol{Y} - \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y} \mid \boldsymbol{Z}])h_g(\boldsymbol{Z}) \\
&\equiv A_\beta(h_\beta) + A_g(h_g).
\end{aligned}
$$

The tangent space $\dot{\mathcal{L}}_{\theta_0}$ (S67) can then be expressed as the range of A:

(S76) $$\dot{\mathcal{L}}_{\theta_0} = A(\mathbb{R} \times \mathcal{H}_g).$$

As discussed in (Van Der Vaart, 1998, Section 25.4), the efficient score for β is

(S77) $$\tilde{S} = A_\beta - \Pi_{\beta,g} A_\beta,$$

where $\Pi_{\beta,g}$ is the orthogonal projection onto the closure $\overline{A_g(\mathcal{H}_g)}$ of the nuisance tangent space $A_g(\mathcal{H}_g)$ in $L^2(\mathcal{L}_{\theta_0})$. In other words,

$$(S78) \qquad \Pi_{\beta,g}A_\beta = \underset{W \in \overline{A_g(\mathcal{H}_g)}}{\arg\min} \|A_\beta - W\|_{L^2(\mathcal{L}_{\theta_0})}.$$

To compute this projection, we first claim that the extended operator $A_g : L^2(\nu) \to L^2(\mathcal{L}_{\theta_0})$ is continuous and that $A_g^*A_g$ is continuously invertible. To verify continuity of A_g, note that for $h_g \in L^2(\nu)$ we have

$$
\begin{aligned}
\|A_g(h_g)\|^2_{L^2(\mathcal{L}_{\theta_0})} &= \mathbb{E}_{\mathcal{L}_{\theta_0}}\left[\left((\boldsymbol{Y} - \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y} \mid \boldsymbol{Z}])h_g(\boldsymbol{Z})\right)^2\right] \\
&= \mathbb{E}_{\mathcal{L}_{\theta_0}}[(\boldsymbol{Y} - \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y} \mid \boldsymbol{Z}])^2 h_g^2(\boldsymbol{Z})] \\
(S79) \qquad &= \mathbb{E}_{\mathcal{L}_{\theta_0}}[\ddot{\psi}(g(\boldsymbol{Z}))h_g^2(\boldsymbol{Z})] \\
&\le C\mathbb{E}_{\mathcal{L}_{\theta_0}}[h_g^2(\boldsymbol{Z})] \\
&= C\|h_g\|^2_{L^2(\nu)}.
\end{aligned}
$$

Next, we derive the adjoint operator $A_g^* : L^2(\mathcal{L}_{\theta_0}) \to L^2(\nu)$. For a random variable $W \in L^2(\mathcal{L}_{\theta_0})$, we have

$$
\begin{aligned}
\langle W, A_g h_g\rangle_{L^2(\mathcal{L}_{\theta_0})} &= \mathbb{E}_{\mathcal{L}_{\theta_0}}\left[W(\boldsymbol{Y} - \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y} \mid \boldsymbol{Z}])h_g(\boldsymbol{Z})\right] \\
(S80) \qquad &= \mathbb{E}_{\mathcal{L}_{\theta_0}}[\mathbb{E}_{\mathcal{L}_{\theta_0}}[W(\boldsymbol{Y} - \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y} \mid \boldsymbol{Z}]) \mid \boldsymbol{Z}]h_g(\boldsymbol{Z})] \\
&= \langle \mathbb{E}_{\mathcal{L}_{\theta_0}}[W(\boldsymbol{Y} - \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y} \mid \cdot]) \mid \cdot], h_g\rangle_{L^2(\nu)}.
\end{aligned}
$$

It follows that

$$(S81) \qquad (A_g^*W)(z) = \mathbb{E}_{\mathcal{L}_{\theta_0}}[W(\boldsymbol{Y} - \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y} \mid \boldsymbol{Z} = z]) \mid \boldsymbol{Z} = z].$$

Next we derive that

$$
\begin{aligned}
(A_g^*A_g h_\eta)(z) &= \mathbb{E}_{\mathcal{L}_{\theta_0}}[A_g h_\eta(\boldsymbol{Y} - \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y} \mid \boldsymbol{Z} = z]) \mid \boldsymbol{Z} = z] \\
&= \mathbb{E}_{\mathcal{L}_{\theta_0}}[(\boldsymbol{Y} - \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y} \mid \boldsymbol{Z}])h_g(\boldsymbol{Z})(\boldsymbol{Y} - \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y} \mid \boldsymbol{Z} = z]) \mid \boldsymbol{Z} = z] \\
&= \mathrm{Var}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y} \mid \boldsymbol{Z} = z]h_g(z) \\
&= \ddot{\psi}(g(z))h_g(z).
\end{aligned}
$$

The assumption (S63) implies that $\ddot{\psi}(g_0(\boldsymbol{Z})) \ge c > 0$ almost surely, since either $\ddot{\psi}$ is a nonzero constant or $g_0(\boldsymbol{Z})$ belongs to a compact set almost surely and therefore $\ddot{\psi}(g_0(\boldsymbol{Z}))$ belongs to the range of a positive continuous function applied to a compact set. From this it follows that $S_g^*S_g$ is continuously invertible. Because A_g is a continuous linear operator with continuously invertible $A_g^*A_g$, it follows that $A_g(L^2(\nu))$ is closed and that $A_g(A_g^*A_g)^{-1}A_g^*$ is the orthogonal projection onto this space. Next let us compute the orthogonal projection of the score A_β onto $A_g(L^2(\nu))$. We have

$$
\begin{aligned}
(A_g^*A_\beta)(z) &= \mathbb{E}_{\mathcal{L}_{\theta_0}}[A_\beta(\boldsymbol{Y} - \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y} \mid \boldsymbol{Z} = z]) \mid \boldsymbol{Z} = z] \\
&= \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{X}(\boldsymbol{Y} - \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y} \mid \boldsymbol{Z}])(\boldsymbol{Y} - \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y} \mid \boldsymbol{Z} = z]) \mid \boldsymbol{Z} = z] \\
(S82) \qquad &= \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{X}(\boldsymbol{Y} - \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y} \mid \boldsymbol{Z} = z])^2 \mid \boldsymbol{Z} = z] \\
&= \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{X} \mid \boldsymbol{Z} = z]\ddot{\psi}(g(z)),
\end{aligned}
$$

and therefore

$$A_g(A_g^*A_g)^{-1}A_g^*A_\beta = (\boldsymbol{Y} - \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y} \mid \boldsymbol{Z}])((A_g^*A_g)^{-1}A_g^*A_\beta)(\boldsymbol{Z})$$

(S83)
$$= (\boldsymbol{Y} - \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y} \mid \boldsymbol{Z}])\mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{X} \mid \boldsymbol{Z}]$$

$$= A_g(\mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{X} \mid \cdot]).$$

Therefore, we have

(S84)
$$\underset{W \in A_g(L^2(\nu))}{\arg\min} \|A_\beta - W\|_{L^2(\mathcal{L}_{\theta_0})} = A_g(\mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{X} \mid \cdot]).$$

Since $A_g(L^2(\nu))$ is closed, it follows that $\overline{A_g(\mathcal{H}_g)} \subseteq A_g(L^2(\nu))$. Together with the assumption (S64) that $\mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{X} \mid \cdot] \in \mathcal{H}_g$ and the definition of the effective score as a projection onto $A_g(\mathcal{H}_g)$ (S78), we deduce that

$$\Pi_{\beta,g}A_\beta = \underset{W \in \overline{A_g(\mathcal{H}_g)}}{\arg\min} \|A_\beta - W\|_{L^2(\mathcal{L}_{\theta_0})} = \underset{W \in A_g(L^2(\nu))}{\arg\min} \|A_\beta - W\|_{L^2(\mathcal{L}_{\theta_0})} = A_g(\mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{X} \mid \cdot]).$$

Therefore, the efficient score is

(S85)
$$\tilde{S} = A_\beta - \Pi_{\beta,g}A_\beta$$

$$= \boldsymbol{X}(\boldsymbol{Y} - \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y} \mid \boldsymbol{Z}]) - (\boldsymbol{Y} - \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y} \mid \boldsymbol{Z}])\mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{X} \mid \boldsymbol{Z}]$$

$$= (\boldsymbol{X} - \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{X} \mid \boldsymbol{Z}])(\boldsymbol{Y} - \mathbb{E}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y} \mid \boldsymbol{Z}]),$$

as claimed (S68). From there we find that the efficient information is

(S86)
$$\widetilde{I}_{\theta_0} = \mathrm{Var}_{\mathcal{L}_{\theta_0}}[\tilde{S}] = \mathbb{E}_{\mathcal{L}_{\theta_0}}[\mathrm{Var}_{\mathcal{L}_{\theta_0}}[\boldsymbol{X}|\boldsymbol{Z}]\mathrm{Var}_{\mathcal{L}_{\theta_0}}[\boldsymbol{Y}|\boldsymbol{Z}]] \equiv s^2(\theta_0),$$

also as claimed (S69).

Differentiability of β and efficient influence function. By Van Der Vaart (1998, Lemma 25.25), the differentiability of β at $\mathcal{L}_{\theta_0}$ with respect to the tangent set $\dot{\mathcal{L}}_{\theta_0}$ follows from the quadratic mean differentiability proved above and the assumption that $\widetilde{I}_{\theta_0} = s^2(\theta_0) > 0$ (S62). The same lemma gives the efficient influence function as the ratio of the efficient score and the efficient information. This completes the proof.

APPENDIX F: ADDITIONAL MATERIAL RELATED TO SIMULATIONS.

In this section, we present details about existing robustness simulation setups (Section F.1), investigate the trade-off between using lasso and post-lasso (Section F.2) and present the complete simulation results (Section F.3).

F.1. Simulation setup in literature. Here we provide details about the simulation setups considered in Candès et al. (2018); Liu et al. (2022); Li and Liu (2022).

Liu et al. (2022). In this paper, the authors consider the double high-dimensional linear model. Suppose $\{Z_i, Y_i\}_{i=1}^n$ is $n = 800$ iid data and $Z_i \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_p)$ where $p = 800$ and $\boldsymbol{\Sigma}_p$ is chosen to be AR(1) and the autocorrelation is set to be 0.5. Then they consider $Y_i = Z_i^\top \beta + \epsilon$ where β is vector of dimension p and only $s = 50$ is set to be nonzero with magnitude $\nu = 0.175$ and random sign. They consider two ways to set the nonzero components of β: spacing these non-zero coefficients equally or choosing them to be the first 50 coefficients of β. The authors consider, rather than testing conditional independence, the false discovery rate (FDR) of variable selection.

Candès et al. (2018).   In this paper, the authors consider a bit different setting where $Y|Z$ is now a high-dimensional logistic model. The sample size $n = 800$ and $Z_i \sim N(\mathbf{0}, \mathbf{\Sigma}_p)$ where $p = 1500$ and $\mathbf{\Sigma}_p$ is chosen to be AR(1) and the autocorrelation is set to be 0.3. After the sampling, the design matrix is centered and every column is normalized to have norm 1. Similarly, only $s = 50$ coordinates of β are set to be nonzero and the sign is random whereas the magnitude is set to be $\nu = 20$. They set a randomly-chosen set of $s = 50$ coefficients of β to be nonzero and consider again the FDR control.

Li and Liu (2022).   In this paper, a similar setting is considered while Z is a data matrix with row $n = 250$ and column $p = 500$ where each row is sampled from $N(0, \mathbf{\Sigma}_p)$ and $\mathbf{\Sigma}_p$ is an AR(1) matrix with autocorrelation 0.5. However, a crucial difference in this paper is the way to set $\mathbb{E}(X|Z)$ and $\mathbb{E}(Y|Z)$. As for X, it is generated by a linear predictor $X = Z\gamma + \epsilon$ where γ is a p dimensional vector with first $s = 5$ is nonzero and the other coordinate remains zero and ϵ follows the standard normal distribution. The sign of each coordinate is randomly chosen and the magnitude is set to be $\nu = 0.3$. As for Y, we set $\beta = \gamma$ and $Y = Z\beta + \xi$ where ξ follows standard normal distribution such that ξ is independent of ϵ. We can see that X and Y support on the same subset of Z so that the marginal association between $\boldsymbol{X}$ and $\boldsymbol{Y}$ is much larger than that in first two simulation designs.

F.2. Comparing the lasso and post-lasso estimation methods.   Compare to the lasso estimation method for $\mu_{n,x}$ and $\mu_{n,y}$, the post-lasso estimation method results in estimates with lower bias but higher variance. This impacts the Type-I error and power of the inferential methods in different ways. For Type-I error, it is only important to have good estimates for $\mathbb{E}[\boldsymbol{X}|\boldsymbol{Z}_\mathcal{A}] \equiv \boldsymbol{Z}_\mathcal{A}^T \beta_\mathcal{A}$ and $\mathbb{E}[\boldsymbol{Y}|\boldsymbol{Z}_\mathcal{A}] \equiv \boldsymbol{Z}_\mathcal{A}^T \gamma_\mathcal{A}$, where $\mathcal{A} \subseteq \{1, \ldots, p\}$ denotes the set of variables active in both $\mathbb{E}[\boldsymbol{X}|\boldsymbol{Z}]$ and $\mathbb{E}[\boldsymbol{Y}|\boldsymbol{Z}]$. Indeed, only the shared active coordinates of $\boldsymbol{Z}$ act as confounders. On the other hand, for power, it is important to have a good estimate for the entire function $\mathbb{E}[\boldsymbol{Y}|\boldsymbol{Z}] = \boldsymbol{Z}^T \gamma$ (Katsevich and Ramdas, 2022). Therefore, we examine the mean-squared estimation error for $\boldsymbol{Z}_\mathcal{A}^T \beta_\mathcal{A}$ and $\boldsymbol{Z}_\mathcal{A}^T \gamma_\mathcal{A}$ in one of our null simulation settings as well as the mean-squared estimation error for $\boldsymbol{Z}^T \beta$ and $\boldsymbol{Z}^T \gamma$ in one of our alternative simulation settings (Figure 1). We find that the post-lasso does a better job estimating the shared active coefficients in the null setting, so the reduced bias in estimating these shared coefficients outweighs the increased variance. On the other hand, the lasso does a better job estimating the entire set of coefficients, so in this case the increased variance outweighs the reduced bias. This explains why the post-lasso-based methods have improved Type-I error control but worse power than the lasso-based methods.

F.3. Additional simulation results.   Figures 2-9 present the complete simulation results across the null and alternative, Gaussian and binary, and supervised and unsupervised settings. All the standard errors are less than 0.026.
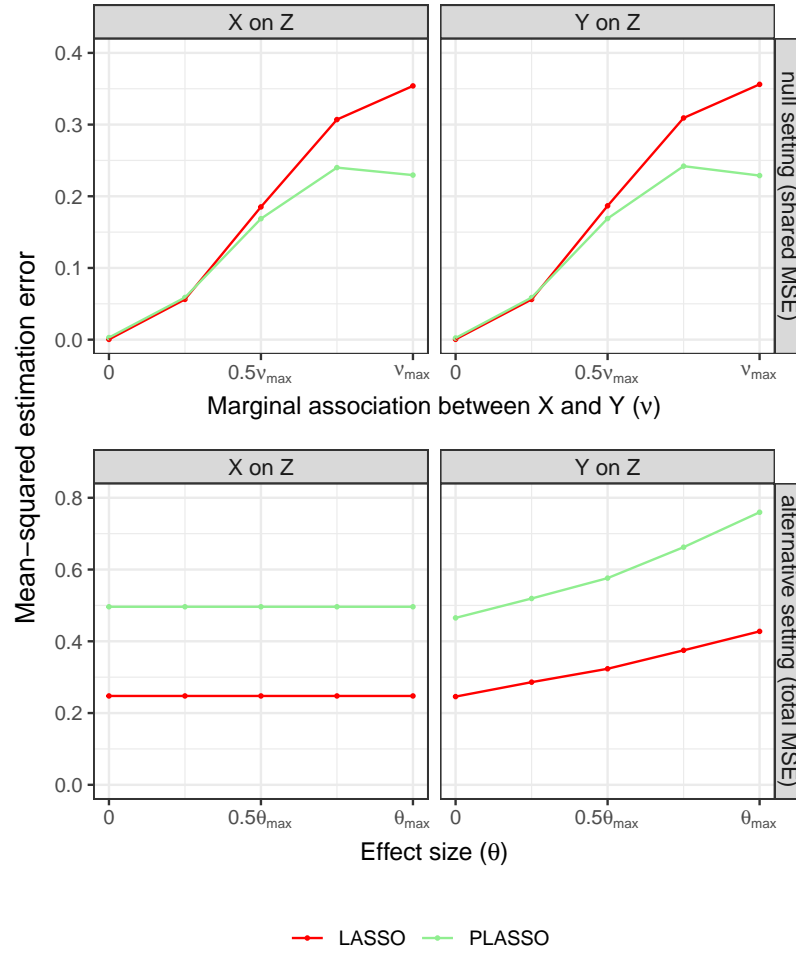
Fig 1: MSE on shared variables and total variables: first column displays the MSE of lasso and post-lasso on shared active variables $Z_{\mathcal{A}}$ and the second column displays the MSE of lasso and post-lasso on total variables. All the experiments are carried out with GCM statistic with $n = 100, d = 400, s = 5, \rho = 0.4$. All the standard errors are less than 0.014.
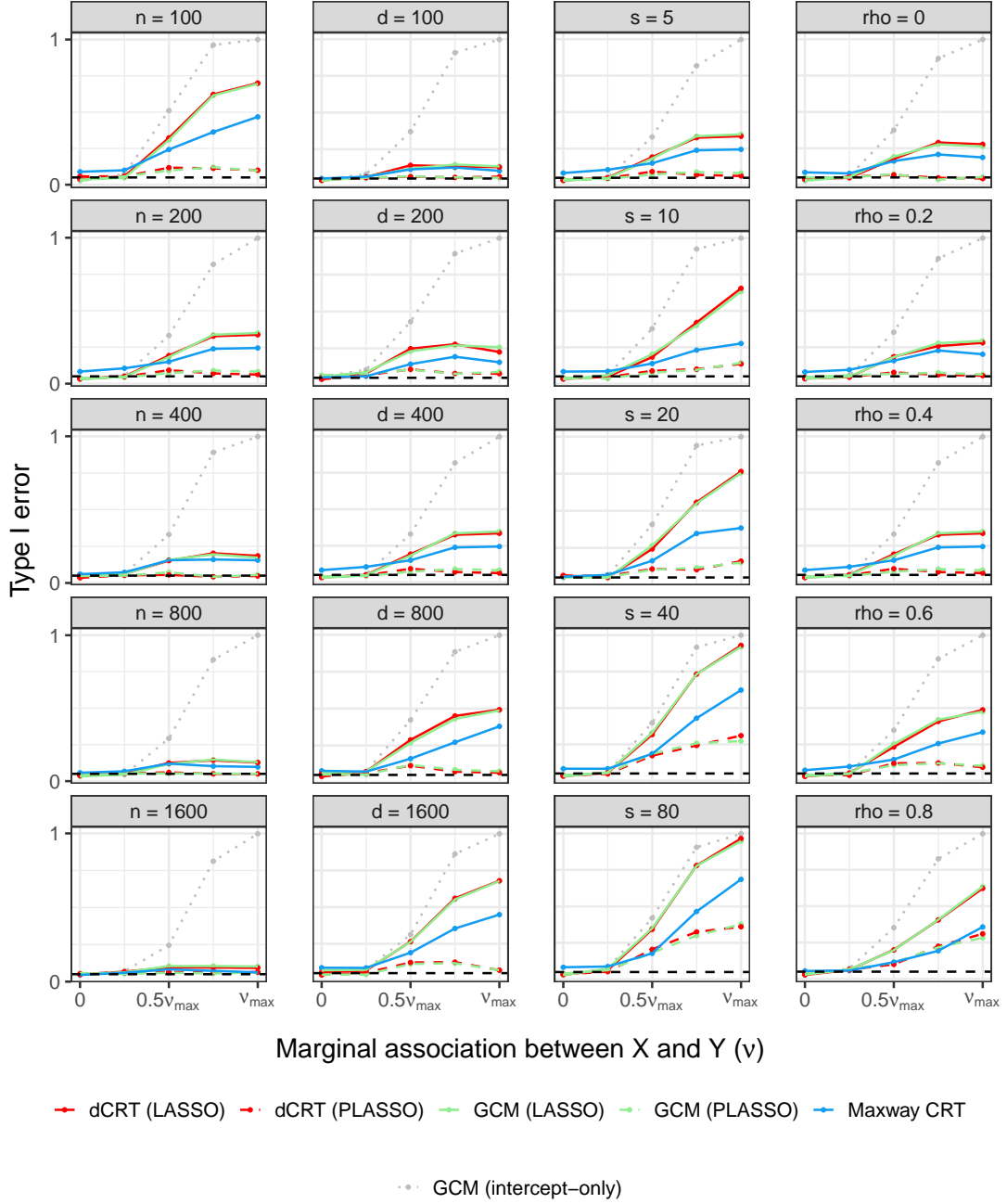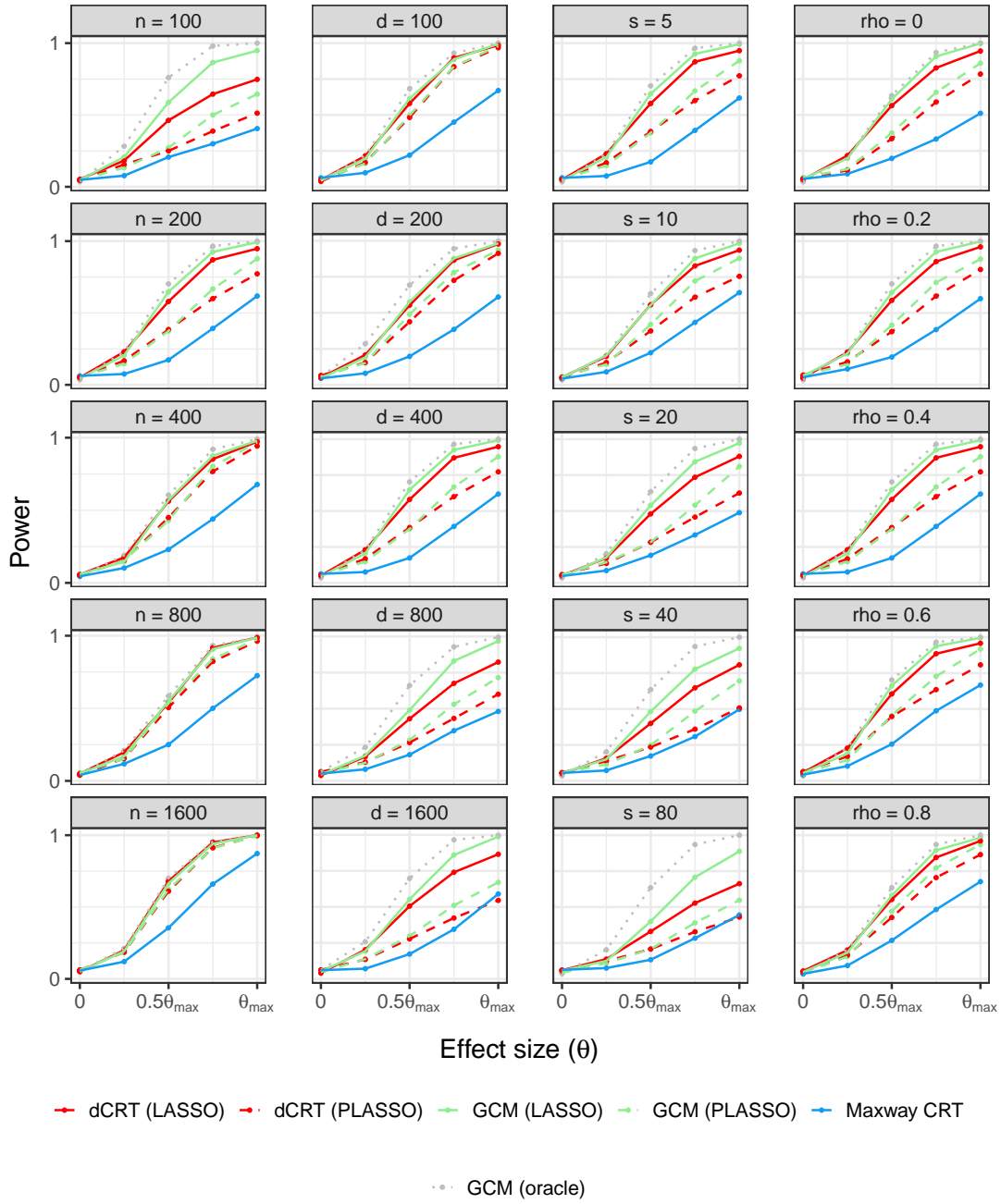
Fig 2: Type-I error in the Gaussian supervised setting.
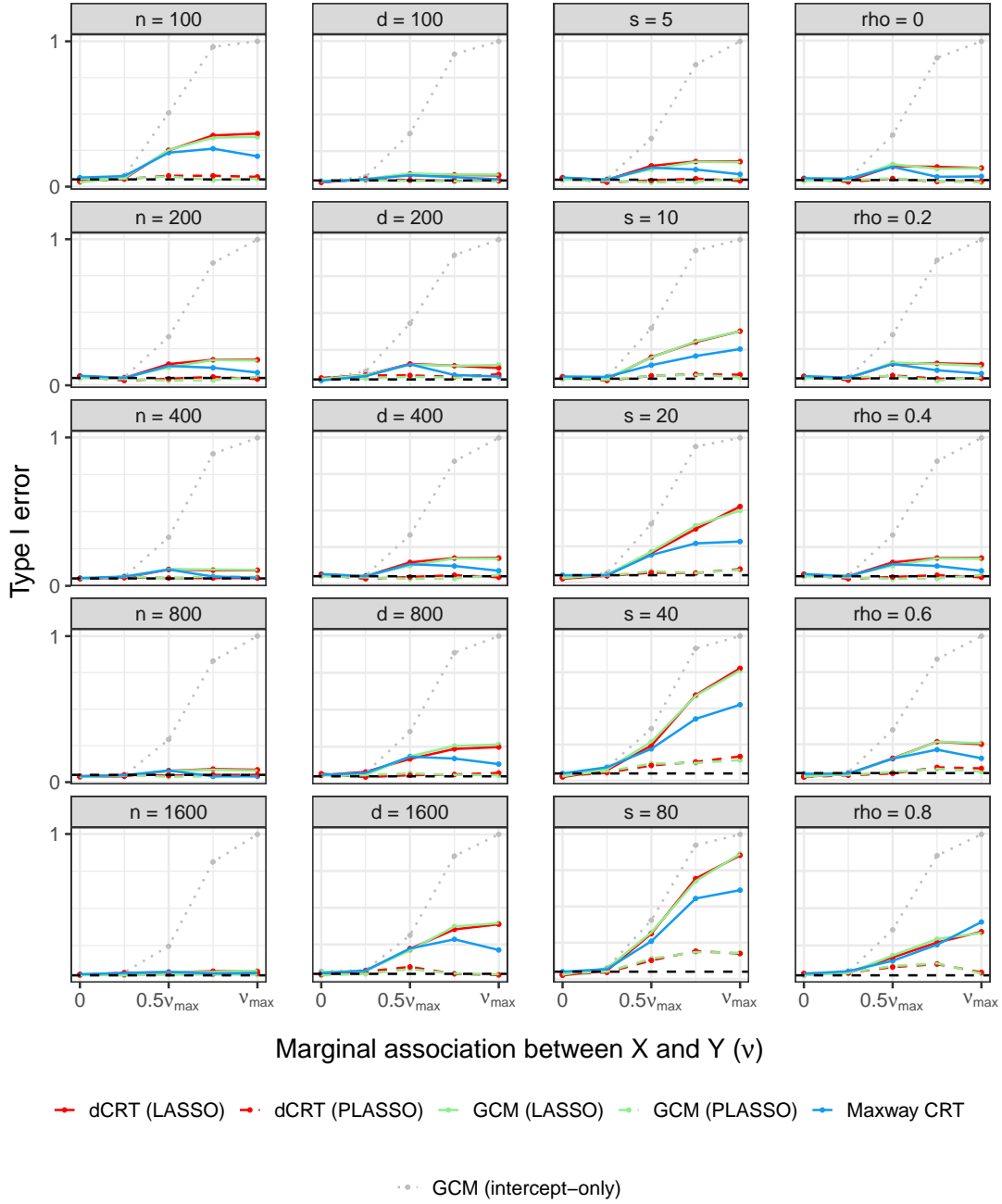
Fig 3: Power in the Gaussian supervised setting.

Fig 4: Type-I error in the Gaussian semi-supervised setting.
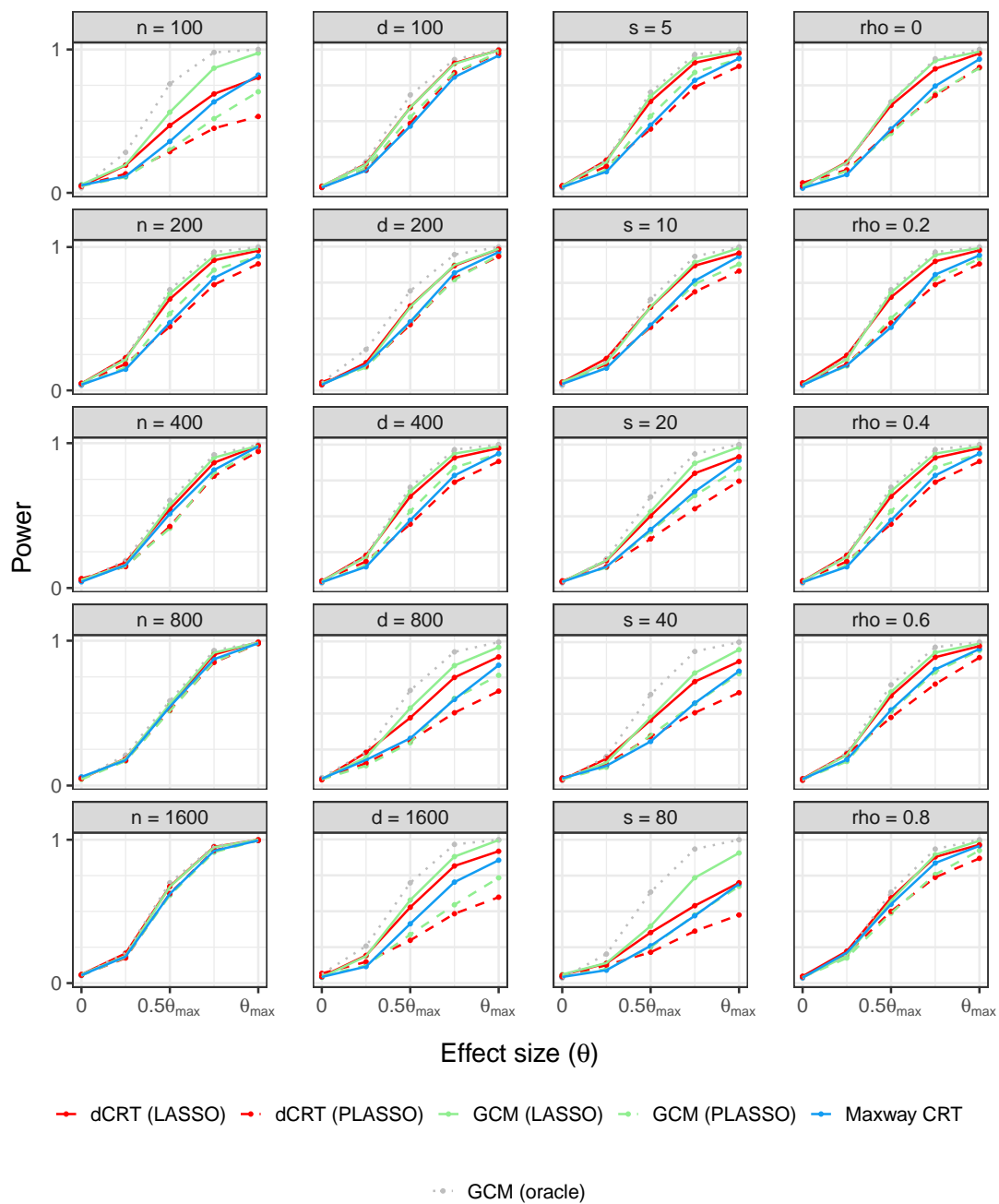
Fig 5: Power in the Gaussian semi-supervised setting.
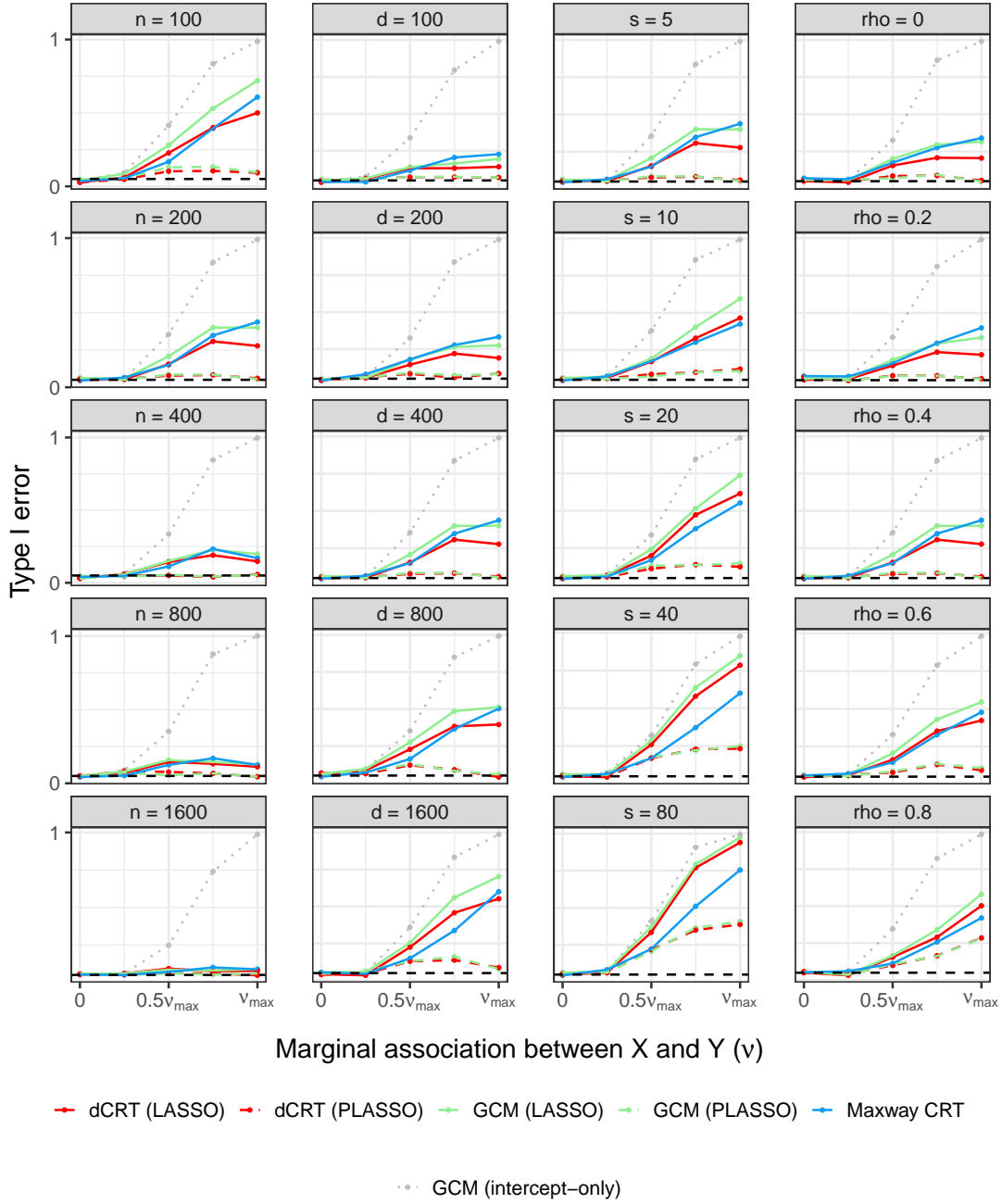
Fig 6: Type-I error in the binary supervised setting.
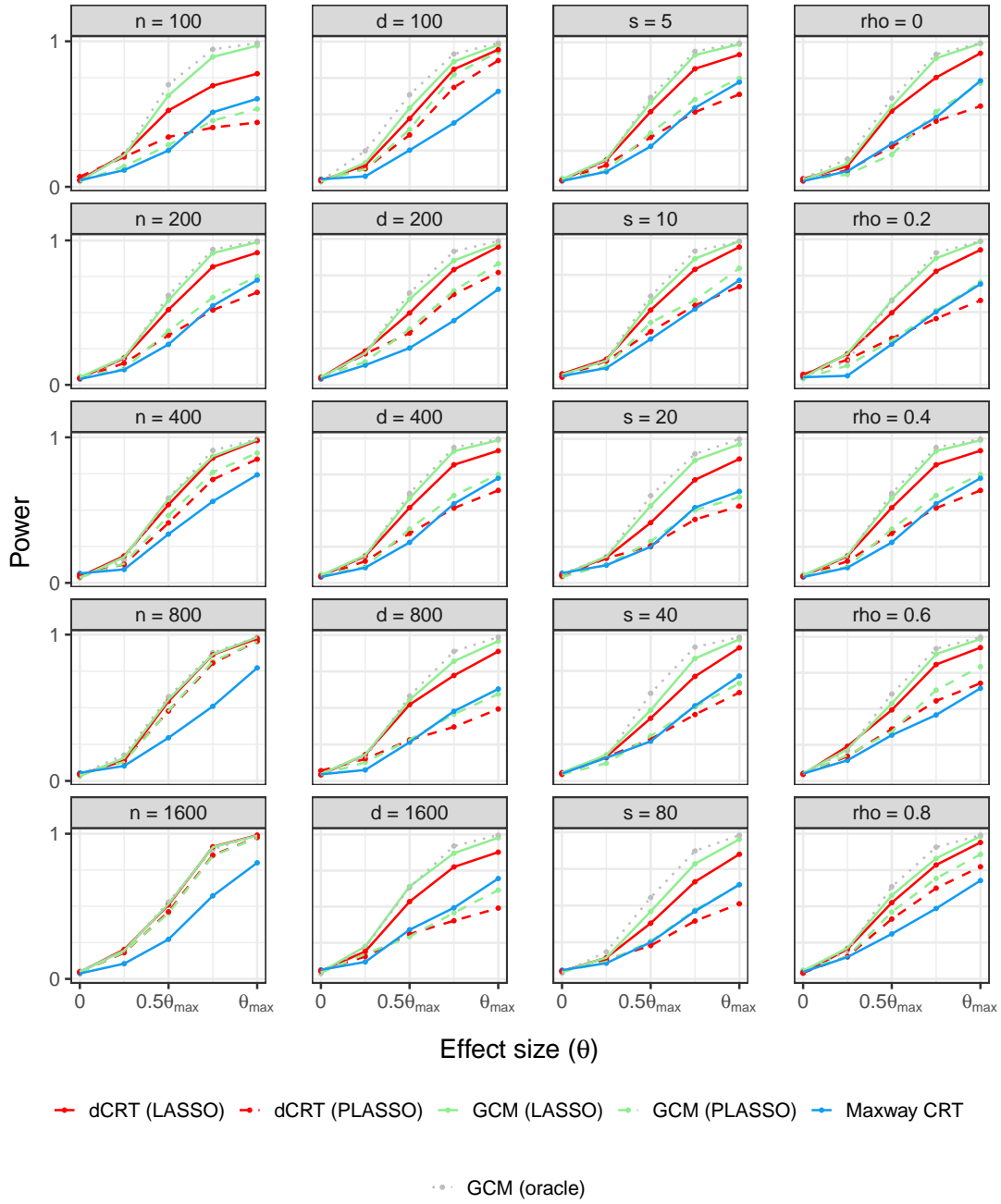
Fig 7: Power in the binary supervised setting.
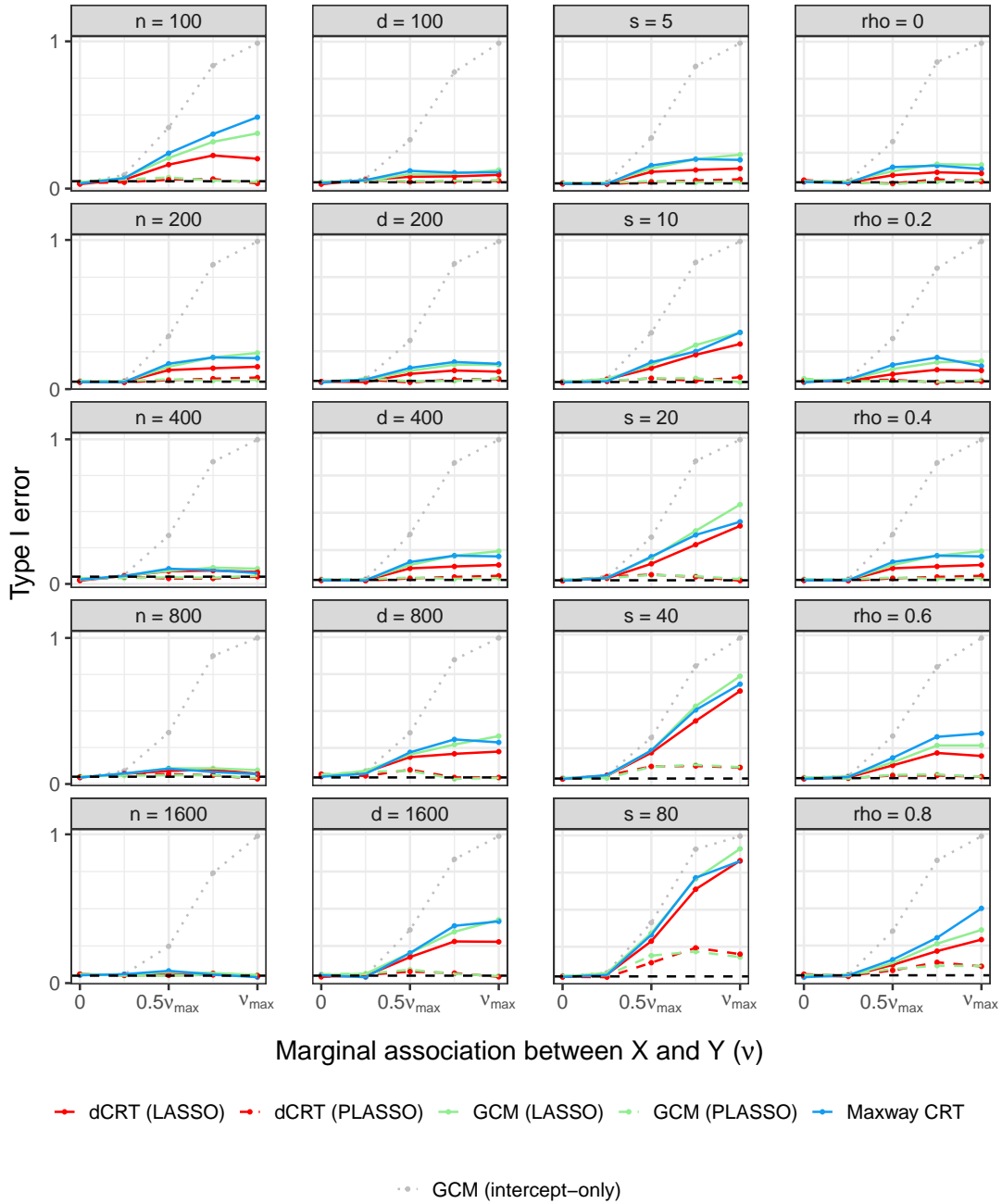
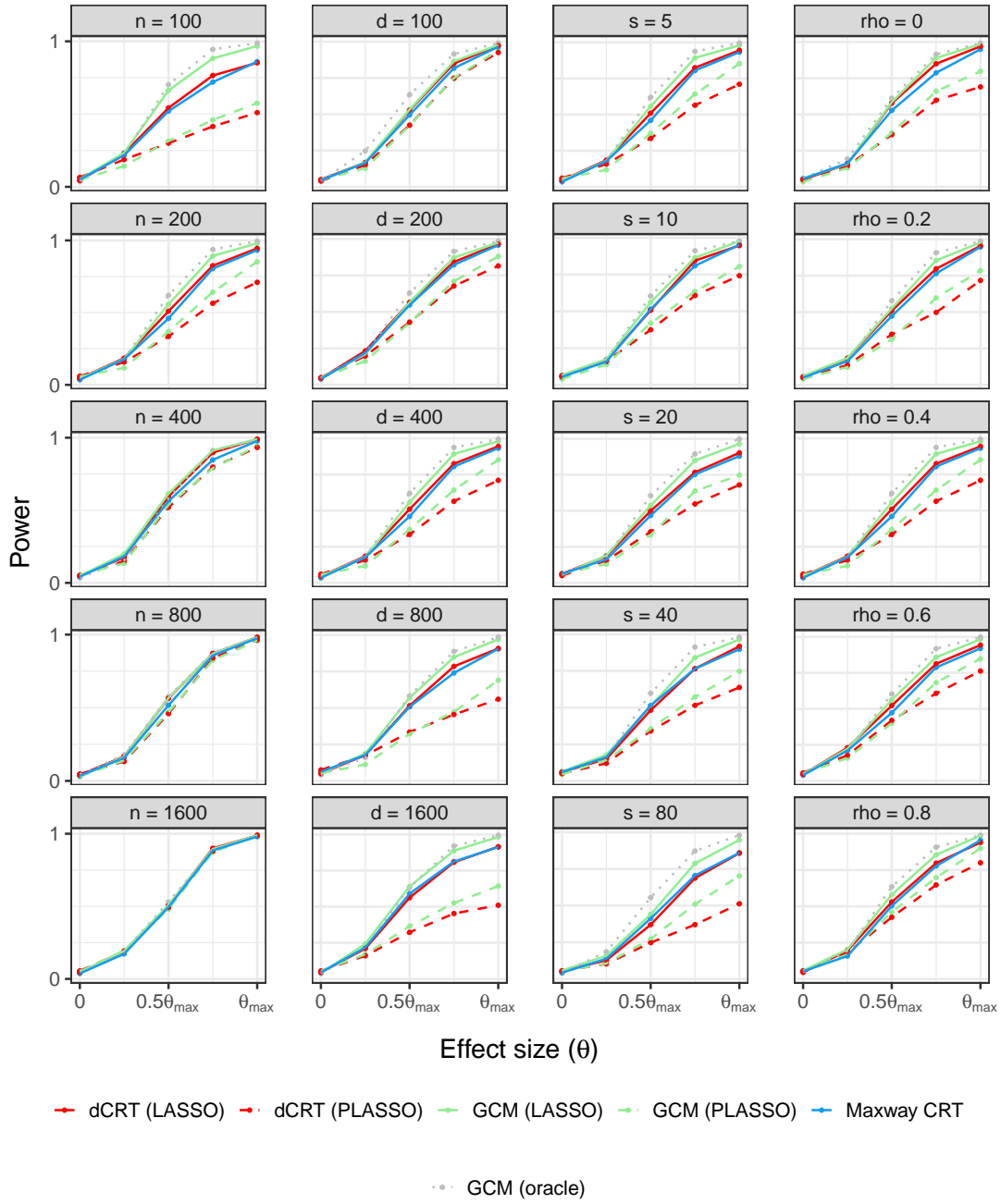Fig 8: Type-I error in the binary semi-supervised setting.

Fig 9: Power in the binary semi-supervised setting.

APPENDIX G: PROOFS OF CONDITIONAL CONVERGENCE RESULTS

In this section, we present the proofs of the conditional convergence results from Appendix B. We proceed by stating and proving necessary lemmas in Section G.1 and then proving the convergence results themselves in Section G.2.

G.1. Auxiliary lemmas. First we state a few known results for the reader's convenience.

LEMMA 10 (Durrett (2010), Theorem 2.3.2). *A sequence of random variables W_n converges to a limit W in probability if and only if every subsequence of W_n has a further subsequence that converges to W almost surely.*

LEMMA 11 (Conditional Markov inequality, Davidson (2003), Theorem 10.17). *Let W be a random variable and let $\mathcal{F}$ be a σ-algebra. If for some $q > 0$ we have $\mathbb{E}[|W|^q] < \infty$, then for any ϵ we have*

$$\mathbb{P}(|W| \geq \epsilon | \mathcal{F}) \leq \frac{\mathbb{E}[|W|^q | \mathcal{F}]}{\epsilon^q} \quad \text{almost surely.}$$

LEMMA 12 (Conditional Hölder inequality, Swanson (2019), Theorem 6.60). *Let W_1 and W_2 be random variables and let $\mathcal{F}$ be a σ-algebra. If for some $q_1, q_2 \in (1, \infty)$ with $\frac{1}{q_1} + \frac{1}{q_2} = 1$ we have $\mathbb{E}[|W_1|^{q_1}], \mathbb{E}[|W_2|^{q_2}] < \infty$, then*

$$\mathbb{E}[|W_1 W_2| \mid \mathcal{F}] \leq (\mathbb{E}[|W_1|^{q_1} \mid \mathcal{F}])^{1/q_1} (\mathbb{E}[|W_2|^{q_2} \mid \mathcal{F}])^{1/q_2} \quad \text{almost surely.}$$

LEMMA 13 (Romano and Lehmann (2005), Lemma 11.2.1). *Suppose $W_n \xrightarrow{d} W$. If for given $\alpha \in (0, 1)$ the CDF of W is continuous and strictly increasing at $\mathbb{Q}_\alpha[W]$, then*

$$\mathbb{Q}_\alpha[W_n] \to \mathbb{Q}_\alpha[W].$$

Next we establish that, without of loss of generality, all random variables, σ-algebras, and conditional expectations in a triangular array may be viewed as being defined on a common probability space.

LEMMA 14 (Embedding into a single probability space). *Consider a sequence of probability spaces $\{(\mathbb{P}_n, \Omega_n, \mathcal{G}_n), n \geq 1\}$. For each n, let $\{W_{i,n}\}_{i \geq 1}$ be a collection of integrable random variables defined on $(\mathbb{P}_n, \Omega_n, \mathcal{G}_n)$ and let $\mathcal{F}_n \subseteq \mathcal{G}_n$ be a σ-algebra. Then there exists a single probability space $(\widetilde{\mathbb{P}}, \widetilde{\Omega}, \widetilde{\mathcal{G}})$, random variables $\{\widetilde{W}_{i,n}\}_{i,n \geq 1}$ on $(\widetilde{\mathbb{P}}, \widetilde{\Omega}, \widetilde{\mathcal{G}})$, and σ-fields $\widetilde{\mathcal{F}}_n \subseteq \widetilde{\mathcal{G}}$ for $n \geq 1$, such that for each n, the joint distribution of $(\{W_{i,n}\}_{i \geq 1}, \{\mathbb{E}[W_{i,n} \mid \mathcal{F}_n]\}_{i \geq 1})$ on $(\mathbb{P}_n, \Omega_n, \mathcal{G}_n)$ coincides with that of $(\{\widetilde{W}_{i,n}\}_{i \geq 1}, \{\mathbb{E}[\widetilde{W}_{i,n} \mid \widetilde{\mathcal{F}}_n]\}_{i \geq 1})$ on $(\widetilde{\mathbb{P}}, \widetilde{\Omega}, \widetilde{\mathcal{G}})$.*

PROOF. Define the Cartesian product $\widetilde{\Omega} \equiv \prod_{n=1}^{\infty} \Omega_n$, the σ-algebra $\widetilde{\mathcal{G}}$ generated by measurable cylinders $\prod_{n=1}^{\infty} A_n$ for $A_n \in \mathcal{G}_n$ and $A_n = \mathcal{G}_n$ for all but finitely many n, and the infinite product measure $\widetilde{\mathbb{P}}$ on the measurable space $(\widetilde{\Omega}, \widetilde{\mathcal{G}})$ (Saeki, 1996). On this probability space, define σ-algebras

(S87) $$\widetilde{\mathcal{F}}_n \equiv \{\mathcal{G}_1 \times \cdots \times \mathcal{G}_{n-1} \times A_n \times \mathcal{G}_{n+1} \times \cdots : A_n \in \mathcal{F}_n\}$$

and random variables

(S88) $$\widetilde{W}_{in}(\omega) \equiv W_{in}(\omega_n)$$

for each $i, n \geq 1$. Next we claim that for each $i, n \geq 1$, the random variable

(S89)
$$\mathbb{E}_{\widetilde{\mathbb{P}}}[\widetilde{W}_{in}|\widetilde{\mathcal{F}}_n](\omega) \equiv \mathbb{E}_{\mathbb{P}_n}[W_{in}|\mathcal{F}_n](\omega_n) \quad \text{for each } \omega \in \widetilde{\Omega}$$

is in fact a version of the conditional expectation $\mathbb{E}_{\widetilde{\mathbb{P}}}[\widetilde{W}_{in}|\widetilde{\mathcal{F}}_n]$. Indeed, it suffices to check that for each $A \equiv \mathcal{G}_1 \times \cdots \times \mathcal{G}_{n-1} \times A_n \times \mathcal{G}_{n+1} \times \cdots \in \widetilde{\mathcal{F}}_n$ we have

(S90)
$$\int_A \mathbb{E}_{\widetilde{\mathbb{P}}}[\widetilde{W}_{in}|\widetilde{\mathcal{F}}_n](\omega)d\widetilde{\mathbb{P}}(\omega) \equiv \int_A \mathbb{E}_{\mathbb{P}_n}[W_{in}|\mathcal{F}_n](\omega_n)d\widetilde{\mathbb{P}}(\omega)$$

$$= \int_{\prod_{n' \neq n} \Omega_{n'}} \int_{A_n} \mathbb{E}_{\mathbb{P}_n}[W_{in}|\mathcal{F}_n](\omega_n)d\widetilde{\mathbb{P}}_n(\omega_n)d\prod_{n' \neq n}\mathbb{P}_{n'}(\omega_{n'})$$

$$\equiv \int_{\prod_{n' \neq n} \Omega_{n'}} \int_{A_n} W_{in}(\omega_n)d\widetilde{\mathbb{P}}_n(\omega_n)d\prod_{n' \neq n}\mathbb{P}_{n'}(\omega_{n'})$$

$$= \int_A W_{in}(\omega_n)d\widetilde{\mathbb{P}}(\omega)$$

$$\equiv \int_A \widetilde{W}_{in}(\omega)d\widetilde{\mathbb{P}}(\omega).$$

From the ω-wise embeddings (S88) and (S89), it is easy to verify the claimed equality between the joint distributions on $(\mathbb{P}_n, \Omega_n, \mathcal{G}_n)$ and $(\widetilde{\mathbb{P}}, \widetilde{\Omega}, \widetilde{\mathcal{G}})$. $\qquad \square$

Finally, we state a conditional version of the truncated weak law of large numbers:

LEMMA 15. *For each n, let $W_{in}, 1 \leq i \leq n$ be a set of random variables independent conditionally on $\mathcal{F}_n$. Let $b_n > 0$ with $b_n \to \infty$ and let $\bar{W}_{in} = W_{in}\mathbb{1}(|W_{in}| \leq b_n)$. Suppose that as $n \to \infty$ we have*

1. $\sum_{i=1}^n \mathbb{P}[|W_{in}| > b_n|\mathcal{F}_n] \xrightarrow{p} 0$ *and*
2. $b_n^{-2} \sum_{i=1}^n \mathbb{E}[\bar{W}_{in}^2|\mathcal{F}_n] \xrightarrow{p} 0$.

If we set $S_n \equiv \sum_{i=1}^n W_{in}$ and $a_n \equiv \sum_{i=1}^n \mathbb{E}[\bar{W}_{in}]$ then

$$\frac{S_n - a_n}{b_n} \mid \mathcal{F}_n \xrightarrow{p,p} 0.$$

PROOF. Let $\bar{S}_n \equiv \sum_{i=1}^n \bar{W}_{in}$. We first write

$$\mathbb{P}\left[\left|\frac{S_n - a_n}{b_n}\right| > \epsilon \middle| \mathcal{F}_n\right] \leq \mathbb{P}\left[S_n \neq \bar{S}_n|\mathcal{F}_n\right] + \mathbb{P}\left[\left|\frac{\bar{S}_n - a_n}{b_n}\right| > \epsilon \middle| \mathcal{F}_n\right].$$

To estimate the first term, we note that

$$\mathbb{P}[S_n \neq \bar{S}_n|\mathcal{F}_n] \leq \mathbb{P}[\cup_{i=1}^n \{\bar{W}_{in} \neq W_{in}\}|\mathcal{F}_n] \leq \sum_{i=1}^n \mathbb{P}(|W_{in}| > b_n|\mathcal{F}_n) \xrightarrow{p} 0$$

by the first assumption. For the second term, we note that conditional Markov's inequality (Lemma 11), $a_n = \mathbb{E}[\bar{S}_n|\mathcal{F}_n]$ implies that

$$\mathbb{P}\left[\left|\frac{\bar{S}_n - a_n}{b_n}\right| > \epsilon|\mathcal{F}_n\right] \leq \epsilon^{-2}\mathbb{E}\left[\left|\frac{\bar{S}_n - a_n}{b_n}\right|^2 \middle| \mathcal{F}_n\right]$$

$$= \epsilon^{-2} b_n^{-2} \mathrm{Var}[\bar{S}_n | \mathcal{F}_n]$$

$$= (b_n \epsilon)^{-2} \sum_{i=1}^{n} \mathrm{Var}[\bar{W}_{in} | \mathcal{F}_n]$$

$$\leq (b_n \epsilon)^{-2} \sum_{i=1}^{n} \mathbb{E}\left[\bar{X}_{in}^2 | \mathcal{F}_n\right] \xrightarrow{p} 0,$$

where the convergence in the last line is given by the second assumption. This completes the proof. $\square$

G.2. Proofs of conditional convergence results.

PROOF OF THEOREM 2. This proof generalizes the argument of (Van Der Vaart, 1998, Lemma 2.11) to allow for conditioning. Fix $\epsilon > 0$, and choose an integer $k \geq 2/\epsilon$. Because the CDF of W is continuous, it follows that

$$\mathbb{P}[W \leq \mathbb{Q}_{i/k}[W]] = i/k \quad \text{for each } 0 \leq i \leq k.$$

Fix $t \in \mathbb{R}$, and suppose $\mathbb{Q}_{\frac{i-1}{k}}[W] \leq t \leq \mathbb{Q}_{\frac{i}{k}}[W]$. It follows that

$$\mathbb{P}[W_n \leq \mathbb{Q}_{\frac{i-1}{k}}[W] \mid \mathcal{F}_n] - \frac{i}{k} \leq \mathbb{P}[W_n \leq t \mid \mathcal{F}_n] - \mathbb{P}[W \leq t] \leq \mathbb{P}[W_n \leq \mathbb{Q}_{\frac{i}{k}}[W] \mid \mathcal{F}_n] - \frac{i-1}{k}.$$

Therefore, for all $t \in \mathbb{R}$, we have

$$|\mathbb{P}[W_n \leq t \mid \mathcal{F}_n] - \mathbb{P}[W \leq t]| \leq \sup_{0 \leq i \leq k} \left| \mathbb{P}[W_n \leq \mathbb{Q}_{i/k}[W] \mid \mathcal{F}_n] - \frac{i}{k} \right| + \frac{1}{k}$$

$$= \sup_{0 \leq i \leq k} \left| \mathbb{P}[W_n \leq \mathbb{Q}_{\frac{i}{k}}[W] \mid \mathcal{F}_n] - \mathbb{P}[W \leq \mathbb{Q}_{\frac{i}{k}}[W]] \right| + \frac{1}{k},$$

so that

$$\sup_{t \in \mathbb{R}} |\mathbb{P}[W_n \leq t \mid \mathcal{F}_n] - \mathbb{P}[W \leq t]| \leq \sup_{0 \leq i \leq k} \left| \mathbb{P}[W_n \leq \mathbb{Q}_{\frac{i}{k}}[W] \mid \mathcal{F}_n] - \mathbb{P}[W \leq \mathbb{Q}_{\frac{i}{k}}[W]] \right| + \frac{1}{k}.$$

By assumption, we have

$$(\text{S91}) \qquad \mathbb{P}\left[\left| \mathbb{P}[W_n \leq \mathbb{Q}_{\frac{i}{k}}[W] \mid \mathcal{F}_n] - \mathbb{P}[W \leq \mathbb{Q}_{\frac{i}{k}}[W]] \right| > \frac{1}{k(k+1)} \right] \to 0.$$

Therefore,

$$\mathbb{P}\left[\sup_{t \in \mathbb{R}} |\mathbb{P}[W_n \leq t \mid \mathcal{F}_n] - \mathbb{P}[W \leq t]| > \epsilon \right]$$

$$\leq \mathbb{P}\left[\sup_{t \in \mathbb{R}} |\mathbb{P}[W_n \leq t \mid \mathcal{F}_n] - \mathbb{P}[W \leq t]| > \frac{2}{k} \right]$$

$$\leq \mathbb{P}\left[\sup_{0 \leq i \leq k} \left| \mathbb{P}[W_n \leq \mathbb{Q}_{\frac{i}{k}}[W] \mid \mathcal{F}_n] - \mathbb{P}[W \leq \mathbb{Q}_{\frac{i}{k}}[W]] \right| > \frac{1}{k} \right]$$

$$\leq \sum_{i=0}^{k} \mathbb{P}\left[\left| \mathbb{P}[W_n \leq \mathbb{Q}_{\frac{i}{k}}[W] \mid \mathcal{F}_n] - \mathbb{P}[W \leq \mathbb{Q}_{\frac{i}{k}}[W]] \right| > \frac{1}{k(k+1)} \right]$$

$$\to 0.$$

This completes the proof. $\square$

PROOF OF THEOREM 3. Fix $t \in \mathbb{R}$. Letting $F(t') \equiv \mathbb{P}[W \leq t']$ be the CDF of W, Theorem 2 gives

$$\left| \mathbb{P}\left[W_n \leq \frac{t - b_n}{a_n} \mid \mathcal{F}_n \right] - F\left(\frac{t - b_n}{a_n} \right) \right| \leq \sup_{t' \in \mathbb{R}} |\mathbb{P}[W_n \leq t' | \mathcal{F}_n] - F(t')| \xrightarrow{p} 0.$$

By the continuous mapping theorem, we have $F\left(\frac{t - b_n}{a_n} \right) \xrightarrow{p} F(t)$, so that

$$\mathbb{P}\left[W_n \leq \frac{t - b_n}{a_n} \mid \mathcal{F}_n \right] \xrightarrow{p} F(t).$$

Noting that $\mathbb{P}[a_n \leq 0 | \mathcal{F}_n]$ is a sequence of nonnegative random variables whose expectations converge to zero, it follows that $\mathbb{P}[a_n \leq 0 | \mathcal{F}_n] \xrightarrow{p} 0$ and so

$$\mathbb{P}\left[a_n W_n + b_n \leq t \mid \mathcal{F}_n \right] = \mathbb{P}\left[a_n W_n + b_n \leq t, a_n > 0 \mid \mathcal{F}_n \right] + o_p(1)$$

$$= \mathbb{P}\left[W_n \leq \frac{t - b_n}{a_n}, a_n > 0 \mid \mathcal{F}_n \right] + o_p(1)$$

$$= \mathbb{P}\left[W_n \leq \frac{t - b_n}{a_n} \mid \mathcal{F}_n \right] + o_p(1)$$

$$\xrightarrow{p} F(t),$$

as desired. $\qquad\square$

PROOF OF THEOREM 4. We apply Lemma 15 with $b_n = n$. We first verify the first assumption in Lemma 15 by conditional Markov's inequality (Lemma 11):

$$\sum_{i=1}^{n} \mathbb{P}(|W_{in}| > n | \mathcal{F}_n) \leq \sum_{i=1}^{n} \frac{\mathbb{E}(|W_{in}|^{1+\delta} | \mathcal{F}_n)}{n^{1+\delta}} \xrightarrow{p} 0.$$

For the second condition, we have

$$\frac{1}{n^2} \sum_{i=1}^{n} \mathbb{E}[W_{in}^2 \mathbb{1}(|W_{in}| \leq n) | \mathcal{F}_n] \leq \frac{1}{n^2} \sum_{i=1}^{n} \mathbb{E}[|W_{in}|^{1+\delta} n^{1-\delta} \mathbb{1}(|W_{in}| \leq n)]$$

$$= \frac{1}{n^{1+\delta}} \sum_{i=1}^{n} \mathbb{E}[|W_{in}|^{1+\delta} \mathbb{1}(|W_{in}| \leq n) | \mathcal{F}_n]$$

$$\leq \frac{1}{n^{1+\delta}} \sum_{i=1}^{n} \mathbb{E}[|W_{in}|^{1+\delta} | \mathcal{F}_n] \xrightarrow{p} 0.$$

Therefore, Lemma 15 yields

$$\left. \frac{1}{n} \sum_{i=1}^{n} (W_{in} - \mathbb{E}[W_{in} \mathbb{1}(|W_{in}| \leq n) \mid \mathcal{F}_n]) \right| \mathcal{F}_n \xrightarrow{p,p} 0.$$

By conditional Slutsky (Theorem 3), it now suffices to show that

$$\frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[W_{in} \mathbb{1}(|W_{in}| > n) \mid \mathcal{F}_n \right] \xrightarrow{p} 0.$$

To see this, applying conditional Markov's and Hölder's inequalities (Lemmas 11 and 12, respectively) we obtain

$$
\left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[W_{in} \mathbb{1}(|W_{in}| > n) | \mathcal{F}_n \right] \right| \leq \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[|W_{in}| \mathbb{1}(|W_{in}| > n) | \mathcal{F}_n \right]
$$

$$
\leq \frac{1}{n} \sum_{i=1}^{n} \{\mathbb{E}[|W_{in}|^{1+\delta} | \mathcal{F}_n]\}^{1/(1+\delta)} \{\mathbb{P}[|W_{in}| > n | \mathcal{F}_n]\}^{\delta/(1+\delta)}
$$

$$
\leq \frac{1}{n} \sum_{i=1}^{n} \{\mathbb{E}[|W_{in}|^{1+\delta} | \mathcal{F}_n]\}^{1/(1+\delta)} \left\{ \frac{\mathbb{E}(|W_{in}|^{1+\delta} | \mathcal{F}_n)}{n^{1+\delta}} \right\}^{\delta/(1+\delta)}
$$

$$
= \frac{1}{n^{1+\delta}} \sum_{i=1}^{n} \mathbb{E}[|W_{in}|^{1+\delta} | \mathcal{F}_n]
$$

$$
\xrightarrow{p} 0,
$$

where the last convergence is by assumption. Finally, we verify that the condition (S10) is sufficient for the conditional WLLN assumption (S8) by noting that it implies

$$
\frac{1}{n^{1+\delta}} \sum_{i=1}^{n} \mathbb{E}[|W_{in}|^{1+\delta} \mid \mathcal{F}_n] \leq \frac{\sup_{1 \leq i \leq n} \mathbb{E}[|W_{in}|^{1+\delta} \mid \mathcal{F}_n] n}{n^{1+\delta}} = \frac{\sup_{1 \leq i \leq n} \mathbb{E}[|W_{in}|^{1+\delta} \mid \mathcal{F}_n]}{n^{\delta}} \xrightarrow{p} 0.
$$

This completes the proof. $\qquad\square$

PROOF OF THEOREM 5. Without loss of generality, we assume $\mathbb{E}[W_{in}|\mathcal{F}_n] = 0$ and that all random variables and σ-algebras are defined on a common probability space $(\mathbb{P}, \Omega, \mathcal{G})$ (Lemma 14). Let $\mathcal{B}(\mathbb{R}^n)$ be the Borel σ-algebra on $\mathbb{R}^n$. Let κ_n be a regular conditional distribution of $(W_{1n}, \ldots, W_{nn})$ given $\mathcal{F}_n$ (Klenke, 2017, Theorem 8.37), i.e. a function $\kappa_n : \Omega \times \mathcal{B}(\mathbb{R}^n) \to [0, \infty]$ such that $\omega \mapsto \kappa_n(\omega, B)$ is measurable for each $B \in \mathcal{B}(\mathbb{R}^n)$, $B \mapsto \kappa_n(\omega, B)$ is a σ-finite measure on $\mathbb{R}^n$ for each $\omega \in \Omega$, and

$$
\kappa_n(\omega, B) = \mathbb{P}[(W_{1n}, \ldots, W_{nn}) \in B | \mathcal{F}_n](\omega), \quad \text{for almost all } \omega \in \Omega \text{ and all } B \in \mathcal{B}(\mathbb{R}^n).
$$

For each n and each $\omega \in \Omega$, let $(\widetilde{W}_{1n}(\omega), \ldots, \widetilde{W}_{nn}(\omega))$ be a draw from the measure $\kappa_n(\omega, \cdot)$. By Klenke (2017, Theorem 8.38), we have for each n that

$$
\left(\sum_{i=1}^{n} \mathrm{Var}[\widetilde{W}_{in}(\omega)] \right)^{-(2+\delta)/2} \sum_{i=1}^{n} \mathbb{E}[|\widetilde{W}_{in}(\omega)|^{2+\delta}] \overset{a.s.}{=} \frac{1}{S_n^{2+\delta}(\omega)} \sum_{i=1}^{n} \mathbb{E}[|W_{in}|^{2+\delta} \mid \mathcal{F}_n](\omega).
$$

Now, let $\{n_k\}_{k \geq 1}$ be a subsequence of $\mathbb{N}$. By the conditional Lyapunov assumption (S15) and Lemma 10, there is a further subsequence n_{k_j} such that

(S92)
$$
\frac{1}{S_{n_{k_j}}^{2+\delta}} \sum_{i=1}^{n_{k_j}} \mathbb{E}[|W_{in_{k_j}}|^{2+\delta} \mid \mathcal{F}_{n_{k_j}}] \overset{a.s.}{\to} 0.
$$

Hence, it follows that
(S93)
$$
\left(\sum_{i=1}^{n_{k_j}} \mathrm{Var}[\widetilde{W}_{in_{k_j}}(\omega)] \right)^{-(2+\delta)/2} \sum_{i=1}^{n_{k_j}} \mathbb{E}[|\widetilde{W}_{in_{k_j}}(\omega)|^{2+\delta}] \to 0 \quad \text{for almost every } \omega \in \Omega.
$$

Applying the usual Lyapunov CLT to the triangular array $\{\widetilde{W}_{in_{k_j}}(\omega)\}_{i,n_{k_j}}$, we find that

(S94) $\qquad \left(\sum_{i=1}^{n_{k_j}} \mathrm{Var}[\widetilde{W}_{in_{k_j}}(\omega)]\right)^{-1/2} \sum_{i=1}^{n_{k_j}} \widetilde{W}_{in_{k_j}}(\omega) \xrightarrow{d} N(0,1) \quad \text{for almost every } \omega \in \Omega,$

and therefore that, for each $t \in \mathbb{R}$, we have
(S95)
$$\mathbb{P}\left[\left(\sum_{i=1}^{n_{k_j}} \mathrm{Var}[\widetilde{W}_{in_{k_j}}(\omega)]\right)^{-1/2} \sum_{i=1}^{n_{k_j}} \widetilde{W}_{in_{k_j}}(\omega) \le t\right] \to \Phi(t) \quad \text{for almost every } \omega \in \Omega.$$

Using Klenke (2017, Theorem 8.38) again, it follows that for each $t \in \mathbb{R}$, we have

(S96) $\qquad \mathbb{P}\left[\dfrac{1}{S_{n_{k_j}}} \sum_{i=1}^{n_{k_j}} W_{in_{k_j}} \le t \mid \mathcal{F}_{n_{k_j}}\right] \xrightarrow{a.s.} \Phi(t).$

Applying Lemma 10, it follows that

(S97) $\qquad \mathbb{P}\left[\dfrac{1}{S_n} \sum_{i=1}^{n} W_{in} \le t \mid \mathcal{F}_n\right] \xrightarrow{p} \Phi(t),$

as desired. $\hfill\square$

PROOF OF LEMMA 1. Without loss of generality, we assume that all random variables and σ-algebras are defined on a common probability space $(\mathbb{P}, \Omega, \mathcal{G})$ (Lemma 14). Let $\mathcal{B}(\mathbb{R})$ be the Borel σ-algebra on $\mathbb{R}$. Let κ_n be a regular conditional distribution of W_n given $\mathcal{F}_n$ (Klenke, 2017, Theorem 8.29), i.e. a function $\kappa_n : \Omega \times \mathcal{B}(\mathbb{R}) \to [0, \infty]$ such that $\omega \mapsto \kappa_n(\omega, B)$ is measurable for each $B \in \mathcal{B}(\mathbb{R})$, $B \mapsto \kappa_n(\omega, B)$ is a σ-finite measure on $\mathbb{R}^n$ for each $\omega \in \Omega$, and

$$\kappa_n(\omega, B) = \mathbb{P}[W_n \in B | \mathcal{F}_n](\omega), \quad \text{for almost all } \omega \in \Omega \text{ and all } B \in \mathcal{B}(\mathbb{R}).$$

Now, let $\{n_k\}_{k \ge 1}$ be a subsequence of $\mathbb{N}$. By conditional Polya's theorem (Theorem 2), we have

(S98) $\qquad \sup_{t \in \mathbb{R}} |\mathbb{P}[W_n \le t | \mathcal{F}_n] - \mathbb{P}[W \le t]| \xrightarrow{p} 0.$

Hence, by Lemma 10 there is a further subsequence n_{k_j} such that

(S99) $\qquad \sup_{t \in \mathbb{R}} |\mathbb{P}[W_{n_{k_j}} \le t | \mathcal{F}_{n_{k_j}}](\omega) - \mathbb{P}(W \le t)| \to 0, \text{ for almost all } \omega \in \Omega.$

It follows that

(S100) $\qquad \sup_{t \in \mathbb{R}} |\kappa_{n_{k_j}}(\omega, (-\infty, t]) - \mathbb{P}(W \le t)| \to 0, \text{ for almost all } \omega \in \Omega,$

i.e.

(S101) $\qquad \kappa_{n_{k_j}}(\omega, \cdot) \xrightarrow{d} W \text{ for almost all } \omega \in \Omega.$

Hence, by Lemma 13, it follows that

(S102) $\qquad \mathbb{Q}_\alpha[\kappa_{n_{k_j}}(\omega, \cdot)] \to \mathbb{Q}_\alpha[W] \text{ for almost all } \omega \in \Omega,$

and therefore

(S103) $\qquad \mathbb{Q}_\alpha[W_{n_{k_j}} | \mathcal{F}_{n_{k_j}}] \xrightarrow{a.s.} \mathbb{Q}_\alpha[W].$

Applying Lemma 10 again, we conclude that

(S104) $\qquad \mathbb{Q}_\alpha[W_n | \mathcal{F}_n] \xrightarrow{p} \mathbb{Q}_\alpha[W],$

as desired. $\hfill\square$

APPENDIX H: FINITE-SAMPLE IMPROVEMENT USING $\widehat{\text{ndCRT}}$ OVER GCM

We proved that the $\widehat{\text{dCRT}}$ and GCM test are asymptotically equivalent, both under the null and under local alternatives (Theorem 2 and Corollary 2). Despite this asymptotic equivalence, the two methods can behave quite differently in finite samples. In particular, the resampling-based $\widehat{\text{dCRT}}$ method can have better finite-sample performance than the asymptotic GCM test. In this appendix, we present a simple numerical example to illustrate this phenomenon, focusing on the Type-I error of the two methods. In order to isolate the effect of the resampling rather than of the test statistic, we perform the comparison using the GCM test statistic for both methods. Therefore, we consider the $\widehat{\text{ndCRT}}$ variant (Appendix A) rather than the $\widehat{\text{dCRT}}$ itself.

Consider the null data-generating model

$$\boldsymbol{Y} \sim \text{Poisson}(\mu_y(\boldsymbol{Z})), \ \boldsymbol{X} \sim \text{Bern}(\mu_x(\boldsymbol{Z}))$$

where $\mu_y(\boldsymbol{Z}) = \exp(\beta_0 + \boldsymbol{Z}\beta_1)$ and $\mu_x(\boldsymbol{Z}) = \text{expit}(\gamma_0 + \boldsymbol{Z}\gamma_1)$ and $\boldsymbol{Z} \sim N(0,1)$. In other words, $\boldsymbol{Y}|\boldsymbol{Z}$ is a Poisson regression model and $\boldsymbol{X}|\boldsymbol{Z}$ is a logistic regression model. We consider the following set of parameters:

$$\gamma_0 = -4, \ \gamma_1 = 1, \ \beta_0 = -3, \ \beta_1 = 1,$$

chosen to make both X and Y sparse. Such a setting occurs commonly in single-cell genomics, and is one where we would expect the convergence of the central limit theorem to be slower. In this simulation, we take the sample size $n = 1000$ and employ the usual generalized linear model estimates for $\mu_x(\cdot)$ and $\mu_y(\cdot)$. We set $M = 1000$ when applying $\widehat{\text{ndCRT}}$ and the simulation is repeated 1000 times.

Figure 10 shows the results of the simulation. The top-left panel compares the p-values from the two methods across 1000 realizations of the data, showing that the $\widehat{\text{ndCRT}}$ produces well-calibrated p-values while the GCM test produces inflated p-values. Thresholding these p-values as nominal level $\alpha = 0.05$, we find that the $\widehat{\text{ndCRT}}$ method controls the Type-I error at the nominal level, while the GCM test does not (top-right panel). The reason for the miscalibration of the GCM test in this setting is that the null distribution of the test statistic has not yet converged to normality (bottom-left panel). By contrast, the resampling distribution of the GCM statistic (employed by the $\widehat{\text{ndCRT}}$ to obtain the critical value) aligns well with its null distribution (bottom-right panel), explaining the improved calibration of the $\widehat{\text{ndCRT}}$ method. Finally, the bottom-right panel shows the empirical CDF of the test statistics, which illustrates the difference in the null distributions of the two methods.
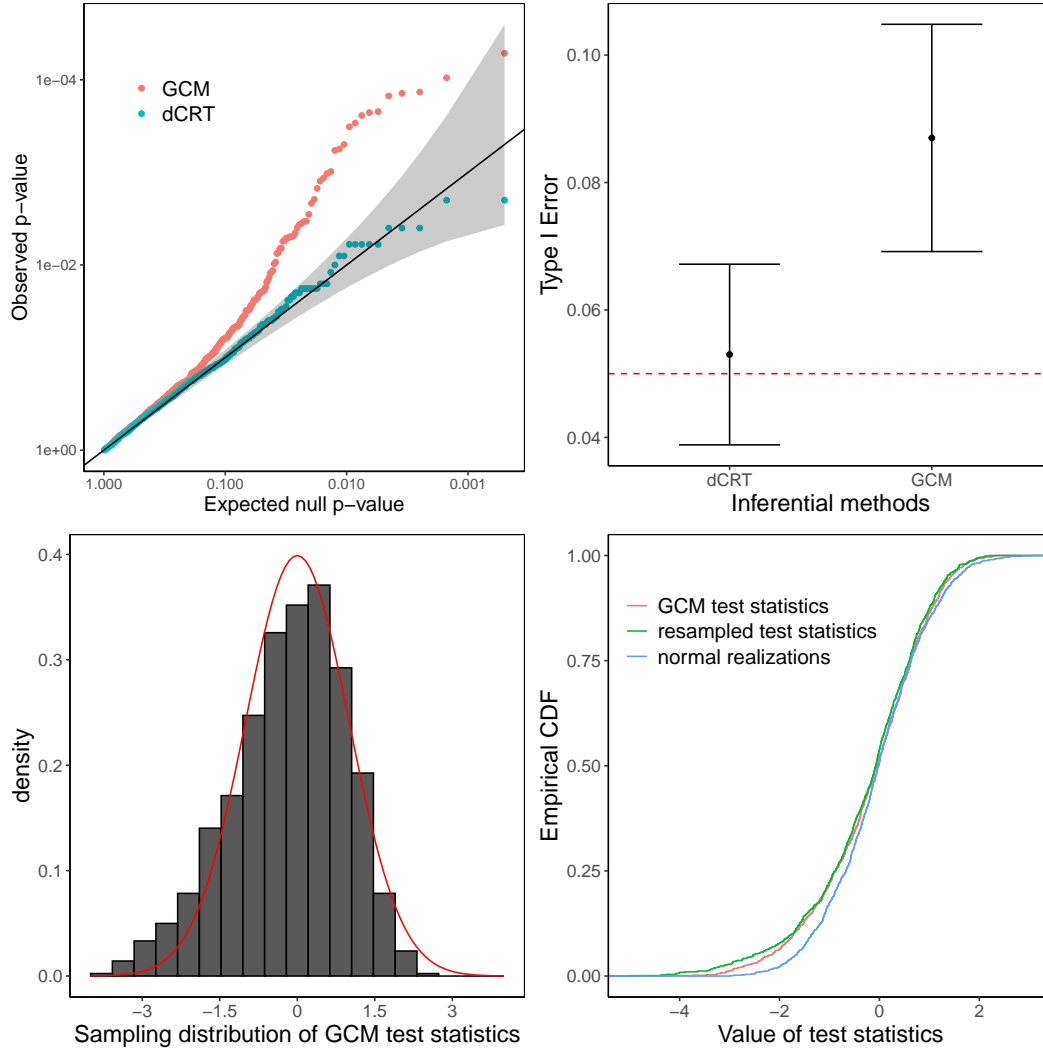
Fig 10: Comparing the calibration of the GCM and $\widehat{\text{ndCRT}}$ methods in a sparse regime. Top left: QQ-plots of the p-values produced by the two methods in a null simulation setting. Top right: Type-I error of the two methods at $\alpha = 0.05$. Bottom left: The true null distribution of the GCM statistic (gray histogram), as well as the standard normal density (red curve). Bottom right: Comparisons of the empirical CDFs of the GCM test statistics, the $\widehat{\text{ndCRT}}$ resamples, and the standard normal distribution.

APPENDIX I: IMPACT OF THE ASYMMETRY OF $\widehat{\text{dCRT}}$

As pointed out by a referee, the $\widehat{\text{dCRT}}$ is asymmetric in X and Y while the conditional independence null hypothesis is symmetric in $\boldsymbol{X}$ and $\boldsymbol{Y}$. One may obtain two different variants of the $\widehat{\text{dCRT}}$ by swapping the roles of X and Y. In this section, we investigate the impact of this asymmetry on the performance of the $\widehat{\text{dCRT}}$ test.

Intuition suggests that the $\widehat{\text{dCRT}}$ will have better performance if resampling is done based on the conditional distribution $\mathcal{L}_n(\boldsymbol{X} \mid \boldsymbol{Z})$ or $\mathcal{L}_n(\boldsymbol{Y} \mid \boldsymbol{Z})$ that is estimated more accurately. To verify this intuition, we consider a simulation setting where the error in estimating $\mathcal{L}_n(\boldsymbol{X} \mid \boldsymbol{Z})$ is less than that in estimating $\mathcal{L}_n(\boldsymbol{Y} \mid \boldsymbol{Z})$. In particular, we a null setting where $\mathcal{L}_n(\boldsymbol{X} \mid \boldsymbol{Z})$ and $\mathcal{L}_n(\boldsymbol{Y} \mid \boldsymbol{Z})$ are Poisson regression models, where the coefficients in $\mathcal{L}_n(\boldsymbol{X} \mid \boldsymbol{Z})$ are smaller than those in $\mathcal{L}_n(\boldsymbol{Y} \mid \boldsymbol{Z})$:

$$\boldsymbol{X}|\boldsymbol{Z} \sim \text{Poisson}(\exp(-1 + \boldsymbol{Z}^\top \beta)), \ \boldsymbol{Y}|\boldsymbol{Z} \sim \text{Poisson}(\exp(-1 + \boldsymbol{Z}^\top \gamma)), \ \boldsymbol{Z} \sim N(\boldsymbol{0}, \boldsymbol{I}_p),$$

where

$$\beta = 0.1 \times (1, 1, \ldots, 1)^\top, \ \gamma = 0.6 \times (1, 1, \ldots, 1)^\top, \ n = 200, \ p \in \{10, 12, \ldots, 18\}.$$

We expect that the coefficient estimates for β will have lower variance than those for γ, due to the small magnitude of the former and the heteroskedastic nature of the Poisson distribution. Furthermore, the less accurate estimation of γ will be amplified on the mean scale by the exponential inverse link function.

Given this data-generating distribution, we apply the $\widehat{\text{dCRT}}$ using maximum likelihood estimation for $\mathcal{L}_n(\boldsymbol{X} \mid \boldsymbol{Z})$ and $\mathcal{L}_n(\boldsymbol{Y} \mid \boldsymbol{Z})$, considering resampling based on $\widehat{\mathcal{L}}_n(\boldsymbol{X} \mid \boldsymbol{Z})$ or $\widehat{\mathcal{L}}_n(\boldsymbol{Y} \mid \boldsymbol{Z})$. We estimate the Type-I errors of each method based on 5000 simulation replicates (Figure 11). We see that resampling from $\widehat{\mathcal{L}}_n(\boldsymbol{Y}|\boldsymbol{Z})$ leads to a very conservative test (blue curve), while resampling from $\widehat{\mathcal{L}}_n(\boldsymbol{X}|\boldsymbol{Z})$ gives a test controlling Type-I error at roughly the nominal level (red curve). This is consistent with our intuition that the $\widehat{\text{dCRT}}$ will have better performance if resampling is done based on the conditional distribution $\mathcal{L}_n(\boldsymbol{X} \mid \boldsymbol{Z})$ or $\mathcal{L}_n(\boldsymbol{Y} \mid \boldsymbol{Z})$ that is estimated more accurately.

To further build intuition, we compare the resampling distributions for the $\widehat{\text{dCRT}}$ based on resampling from $\widehat{\mathcal{L}}_n(\boldsymbol{X}|\boldsymbol{Z})$ and $\widehat{\mathcal{L}}_n(\boldsymbol{Y}|\boldsymbol{Z})$ to the sampling distribution of the $\widehat{\text{dCRT}}$ test statistic (Figure 12). We find that the resampling distribution of the test statistic is more aligned with the sampling distribution when resampling from $\widehat{\mathcal{L}}_n(\boldsymbol{X}|\boldsymbol{Z})$. On the other hand, the resampling distribution based on $\widehat{\mathcal{L}}_n(\boldsymbol{Y}|\boldsymbol{Z})$ has much greater variation than the sampling distribution, explaining the conservativeness of the corresponding test.

In summary, we have presented a scenario where resampling from $\widehat{\mathcal{L}}_n(\boldsymbol{X}|\boldsymbol{Z})$ and $\widehat{\mathcal{L}}_n(\boldsymbol{Y}|\boldsymbol{Z})$ can result in different Type-I error behavior, which is due to the different levels of accuracy in these estimates. Resampling from the more accurate distribution can result in better Type-I error performance. This difference exists primarily in finite-sample regimes; as the sample size grows, both resampling strategies will converge to the GCM test under the appropriate assumptions (Theorem 2). Nevertheless, this example shows that care should be taken in deciding whether to resample from $\widehat{\mathcal{L}}_n(\boldsymbol{Y}|\boldsymbol{Z})$ or $\widehat{\mathcal{L}}_n(\boldsymbol{X}|\boldsymbol{Z})$.

REFERENCES

BULINSKI, A. V. (2017). Conditional central limit theorem. *Theory of Probability and its Applications* **61** 613–631.

CANDÈS, E., FAN, Y., JANSON, L. and LV, J. (2018). Panning for gold: 'model-X' knockoffs for high dimensional controlled variable selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80** 551–577.
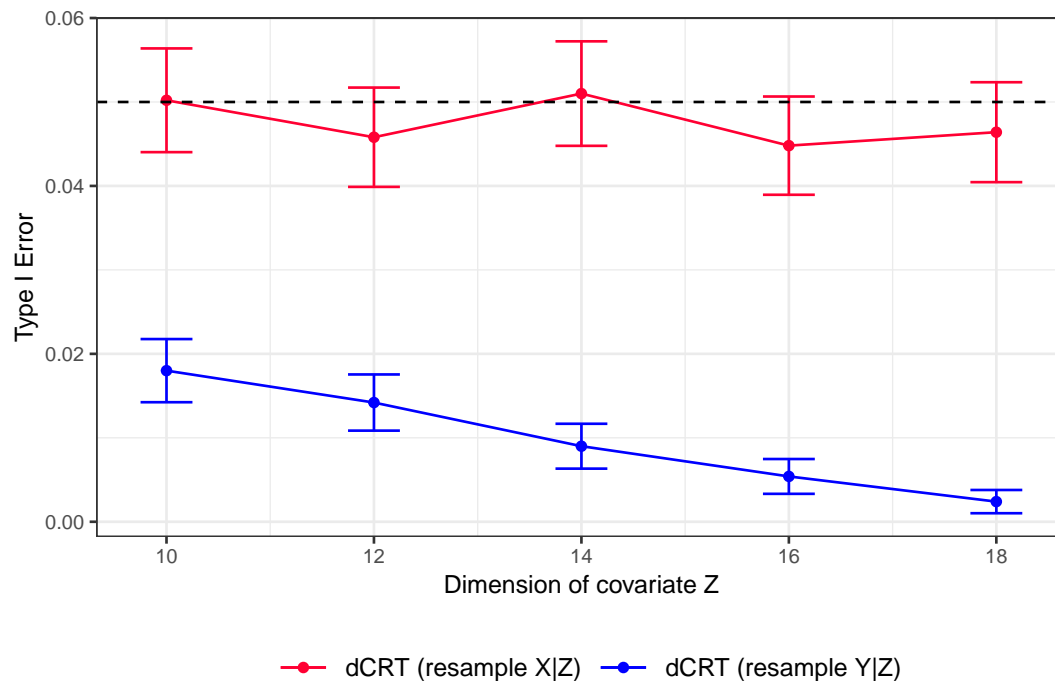
Fig 11: Comparing the Type-I errors of the $\widehat{\mathrm{dCRT}}$ when resampling is done based on $\widehat{\mathcal{L}}_n(\boldsymbol{X}|\boldsymbol{Z})$ or $\widehat{\mathcal{L}}_n(\boldsymbol{Y}|\boldsymbol{Z})$.

CHOI, S., HALL, W. J. and SCHICK, A. (1996). Asymptotically uniformly most powerful tests in parametric and semiparametric models. *Annals of Statistics* **24** 841–861. https://doi.org/10.1214/aos/1032894469

DAVIDSON, J. (2003). Stochastic Limit Theory. *Stochastic Limit Theory*. https://doi.org/10.1093/0198774036.001.0001

DEDECKER, J. and MERLEVEDE, F. (2002). Necessary and sufficient conditions for the conditional central limit theorem. *Annals of Probability* **30** 1044–1081. https://doi.org/10.1137/S0040585X97T98837X

DURRETT, R. (2010). *Probability: Theory and Examples*, 4th ed. Cambridge University Press. https://doi.org/10.1017/9781108591034

GRZENDA, W. and ZIEBA, W. (2008). Conditional central limit theorem. *International Mathematical Forum* **3** 1521–1528.

KATSEVICH, E. and RAMDAS, A. (2022). On the power of conditional independence testing under model-X. *Electronic Journal of Statistics* **16** 6348–6394.

KIM, I. and RAMDAS, A. (2024). Dimension-agnostic inference. *Bernoulli* **30** 683–711.

KLENKE, A. (2017). *Probability theory* **941**.

KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York.

LI, S. and LIU, M. (2022). Maxway CRT: Improving the Robustness of Model-X Inference. *arXiv*.

LIU, M., KATSEVICH, E., JANSON, L. and RAMDAS, A. (2022). Fast and powerful conditional randomization testing via distillation. *Biometrika* **109** 277–293. https://doi.org/10.1093/biomet/asab039

MAJEREK, D., NOWAK, W. and ZIEBA, W. (2005). Conditional strong law of large number. *Int. J. Pure Appl. Math* **20** 143–156.

NOWAK, W. and ZIĘBA, W. (2005). Types of conditional convergence. *Annales Universitatis Mariae Curie-Sklodowska Lublin-Polonia* **59** 97–105.

PRAKASA RAO, B. L. S. (2009). Conditional independence, conditional mixing and conditional association. *Annals of the Institute of Statistical Mathematics* **61** 441–460. https://doi.org/10.1007/s10463-007-0152-2

ROBINS, J. M. and ROTNITZKY, A. (2001). Comment on the Bickel and Kwon article, "Inference for semiparametric models: Some questions and an answer". *Statistica Sinica* **11** 920–936.

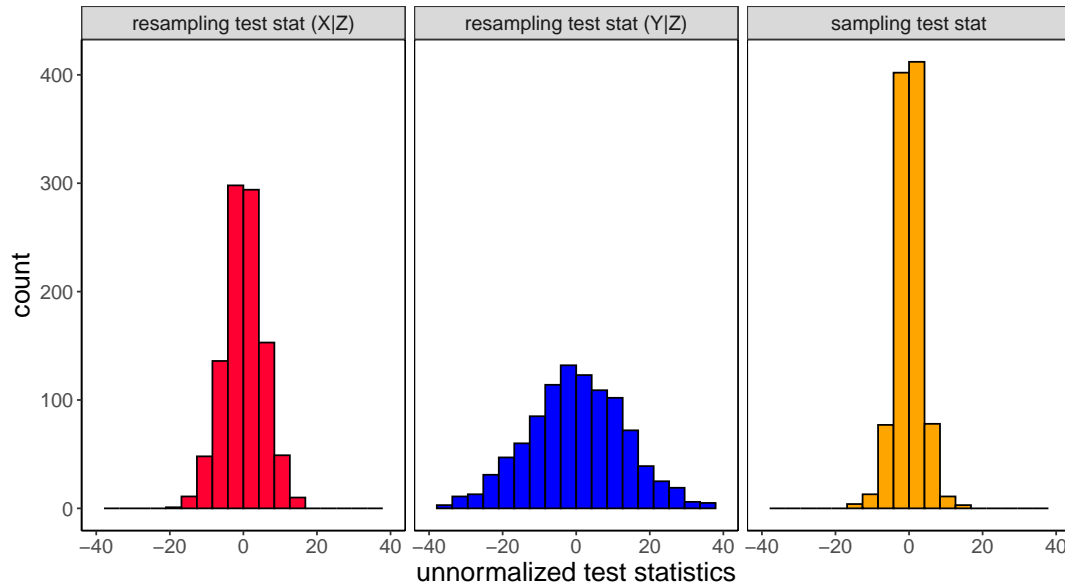ROMANO, J. P. and LEHMANN, E. L. (2005). *Testing Statistical Hypothesis*. https://doi.org/10.2307/2332982

Fig 12: Histograms of the resampling distributions based on resampling from $\widehat{\mathcal{L}}_n(\boldsymbol{X}|\boldsymbol{Z})$ and $\widehat{\mathcal{L}}_n(\boldsymbol{Y}|\boldsymbol{Z})$, as well as the sampling distribution of the same statistics, in the simulation underlying Figure 11.

SAEKI, S. (1996). A Proof of the Existence of Infinite Product Probability Measures. *The American Mathematical Monthly* **103** 682–683. https://doi.org/10.1080/00029890.1996.12004804

SHAH, R. D. and PETERS, J. (2020). The Hardness of Conditional Independence Testing and the Generalised Covariance Measure. *Annals of Statistics* **48** 1514–1538.

SHEKHAR, S., KIM, I. and RAMDAS, A. (2022). A permutation-free kernel two-sample test. *Advances in Neural Information Processing Systems* **35** 18168–18180.

SHEKHAR, S., KIM, I. and RAMDAS, A. (2023). A Permutation-Free Kernel Independence Test. *Journal of Machine Learning Research* **24** 1–68.

SWANSON, J. (2019). Lecture notes on probability theory Technical Report.

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

WAINWRIGHT, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*. https://doi.org/10.1017/9781108627771

WANG, W. and JANSON, L. (2022). A high-dimensional Power Analysis of the Conditional Randomization Test and Knockoffs. *Biometrika* **109** 631–645.

YUAN, D. M., WEI, L. R. and LEI, L. (2014). Conditional central limit theorems for a sequence of conditional independent random variables. *Journal of the Korean Mathematical Society* **51** 1–15. https://doi.org/10.4134/JKMS.2014.51.1.001