
Increasing Effect Sizes of Pairwise Conditional Independence Tests between Random Vectors (Supplementary Material)

Tom Hochsprung¹

Jonas Wahl^{*2,1}

Andreas Gerhardus^{*1}

Urmi Ninad^{*2,1}

Jakob Runge^{1,2}

¹Institute of Data Science, German Aerospace Center, Jena, Germany

²Technische Universität Berlin, Berlin, Germany

A BACKGROUND MATERIAL

A.1 MULTIVARIATE NORMAL DISTRIBUTION

In the following, we recall some basic results about the multivariate normal distribution that are relevant in the context of this work [Anderson, 2003].

Let the joint distribution of $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ be multivariate normal, i.e., $P_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}} = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for some mean vector $\boldsymbol{\mu}$ and some positive definite covariance matrix $\boldsymbol{\Sigma}$. It is a known fact that the multivariate normal distribution is closed under conditioning and marginalization, that means that if $P_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}$ is multivariate normal, then $P_{\mathbf{X}, \mathbf{Y} | \mathbf{Z}=\mathbf{z}}$, $P_{\mathbf{X} | \mathbf{Z}=\mathbf{z}}$, $P_{\mathbf{Y} | \mathbf{Z}=\mathbf{z}}$, $P_{\mathbf{X}}$, $P_{\mathbf{Y}}$ and $P_{\mathbf{Z}}$ are also (multivariate) normal. Furthermore, it is known that the partial correlation coefficient encapsulates the entire dependence structure between components of a multivariate normal random vector; in particular, $X_i \perp\!\!\!\perp Y_j | \mathbf{Z}$ if and only if $\rho_{X_i Y_j | \mathbf{Z}} = 0$.

Conditional independence is also encoded in the covariance matrices $\boldsymbol{\Sigma}^{\mathbf{z}}$ corresponding to the distributions $P_{\mathbf{X}, \mathbf{Y} | \mathbf{Z}=\mathbf{z}}$. The fact $X_i \perp\!\!\!\perp Y_j | \mathbf{Z}$ is equivalent to $\Sigma_{ij}^{\mathbf{z}} = 0$ for all \mathbf{z} . If $X_i \perp\!\!\!\perp Y_j | \mathbf{Z}$ for all $i \in \{1, \dots, d_X\}$ and $j \in \{1, \dots, d_Y\}$, then each $\boldsymbol{\Sigma}^{\mathbf{z}}$ is a block diagonal matrix, which implies that $p_{\mathbf{X}, \mathbf{Y} | \mathbf{Z}=\mathbf{z}}$ factorizes for each \mathbf{z} according to equation (2), which implies $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \mathbf{Z}$.

Thus, if the distribution of $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ is multivariate normal, then statement (1) is indeed equivalent to statement (3). As the multivariate normal distribution is closed under conditioning and marginalization, the above reasoning also applies to arbitrary subvectors \mathbf{X}_A and \mathbf{Y}_B . Thus, if $P_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}$ is multivariate normal, then Assumption 1 is satisfied.

A.2 OTHER SUFFICIENT CONDITIONS FOR ASSUMPTION 1

We now discuss another sufficient condition for Assumption 1 to hold. This sufficient condition arises in the area of graphical modelling.

Following Pearl [2009] and Spirtes et al. [2000], let $G = (V, E)$ be a directed acyclic graph, where V is the set of vertices, and E is the set of directed edges. In slight abuse of notation, we equate the vertex names with the variable names. Now, assume that the joint distribution $P_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}$ is faithful and globally Markov with respect to G . Then, Assumption 1 is satisfied.

To see this fact, note that $X_i \perp\!\!\!\perp Y_j | \mathbf{Z}$ for all $i \in A$ and for all $j \in B$ in conjunction with faithfulness implies that X_i and Y_j are d-separated by \mathbf{Z} for all $i \in A$ and for all $j \in B$. This pairwise d-separation then implies that \mathbf{X}_A and \mathbf{Y}_B are d-separated by \mathbf{Z} , which, in conjunction with the global Markov property, implies that $\mathbf{X}_A \perp\!\!\!\perp \mathbf{Y}_B | \mathbf{Z}$.

*Equal contribution, order chosen uniformly at random.

B MORE NUMERICAL EXPERIMENTS

B.1 DIFFERENT SIGNIFICANCE LEVELS IN FIRST STEP OF SAMPLE SPLITTING ALGORITHM

We do the same simulations as in Section 5, this time, with $\alpha_{pre} = 0.8$ in the first step for the algorithms that use sample splitting. Figure 3 displays the results.

A larger α_{pre} leads to smaller sets Q_i and Q'_j that are learned in the first step. In case of $\Sigma^{(1)}$, where the dependence between \mathbf{X} and \mathbf{Y} is only between X_1 and Y_1 , we would expect a slightly worse performance than for $\alpha_{pre} = 0.5$ because theoretically, the Q_i 's and Q'_j 's can be very large, and by learning smaller subsets, we omit some useful components that can be additionally conditioned out. We see that this theoretical reasoning indeed seems to be true; the results for $\alpha_{pre} = 0.8$ (Figure 3 in the SM) are indeed slightly worse than the results for $\alpha_{pre} = 0.5$ (Figure 2 in the main paper).

If there are more components of \mathbf{X} and \mathbf{Y} that are conditionally dependent (which, for example, is the case for $\Sigma^{(2)}$), then the effect of increasing α_{pre} is harder to foresee. On the one hand, some conditional independencies might be additionally omitted if α_{pre} is large, however, if α_{pre} is too small, wrong components of \mathbf{X} and \mathbf{Y} might be deemed conditionally independent (a Type II error). We again see that the performance for $\Sigma^{(2)}$ is slightly worse for $\alpha_{pre} = 0.8$ (Figure 3 in the SM) than for $\alpha_{pre} = 0.5$ (Figure 2 in the main paper).

B.2 SIMULATION RESULTS FOR THE GENERALIZED COVARIANCE MEASURE

In this section, we emphasize that the results of this paper are not only true for partial correlations and the multivariate normal distribution, but also more generally for other dependence measures and distributions. In particular, we employ the *generalized covariance measure* from Shah and Peters [2020] which is implemented in the R-package *GeneralisedCovarianceMeasure* [Peters and Shah, 2022] and show that we get similar empirical results to the one in Section 5. We again restrict to the case $d_Z = 1$. For the case $d_X = d_Y = 2$, we use a model similar to model (d) in Section 5.2 of Shah and Peters [2020], i.e.,

$$\begin{aligned} X_1 &:= \exp(-Z^2/2) \cdot \sin(Z) + 0.3 \cdot \eta_1, \\ X_2 &:= \exp(-Z^2/2) \cdot \sin(Z) + \tau \cdot X_1 + 0.3 \cdot \eta_2, \\ Y_1 &:= \exp(-Z^2/2) \cdot \sin(Z) + 0.3 \cdot \eta_3, \\ Y_2 &:= \exp(-Z^2/2) \cdot \sin(Z) + \tau \cdot Y_1 + \rho \cdot X_2 + 0.3 \cdot \eta_4, \end{aligned}$$

where Z, η_1, \dots, η_4 are independent standard normal random variables and ρ and τ are real-valued parameters.

For the case $d_X = d_Y = 3$, we look at the model

$$\begin{aligned} X_1 &:= \exp(-Z^2/2) \cdot \sin(Z) + 0.3 \cdot \eta_1, \\ X_2 &:= \exp(-Z^2/2) \cdot \sin(Z) + \tau \cdot X_1 + 0.3 \cdot \eta_2, \\ X_3 &:= \exp(-Z^2/2) \cdot \sin(Z) + \tau \cdot X_2 + 0.3 \cdot \eta_3, \\ Y_1 &:= \exp(-Z^2/2) \cdot \sin(Z) + 0.3 \cdot \eta_4, \\ Y_2 &:= \exp(-Z^2/2) \cdot \sin(Z) + \tau \cdot Y_1 + 0.3 \cdot \eta_5, \\ Y_3 &:= \exp(-Z^2/2) \cdot \sin(Z) + \tau \cdot Y_2 + \rho \cdot X_3 + 0.3 \cdot \eta_6, \end{aligned}$$

where Z, η_1, \dots, η_6 are independent standard normal random variables and ρ and τ are again real-valued parameters.

The parameters τ and ρ have a similar meaning as in Section 5. Roughly speaking, τ characterizes the within-group dependence, and ρ characterizes the between-group dependence. We look at the cases $\tau \in \{0, 1, 2\}$ and $\rho \in \{0, 0.005, \dots, 0.15\}$. Moreover, we consider the sample sizes $n \in \{216, 432, 864\}$.

We compare the same algorithms with the same settings as in Section 5. Again, we employ the Bonferroni method to aggregate univariate test statistics. We again do 100 replications for each of the above mentioned cases and plot the mean rejection rate with one standard error (see Figure 4 in the SM). We observe similar results as in Section 5. If the within-group dependence is relatively high, i.e., $\tau = 2$, the algorithm that assumes a conditional independence oracle (Section 3.2) and the sample splitting algorithms (Section 3.3) perform better; if the within-group dependence is relatively low, i.e., $\tau = 0$, there is no improvement by conditioning out already known independencies. For the case $\tau = 0$, the sample-splitting algorithm

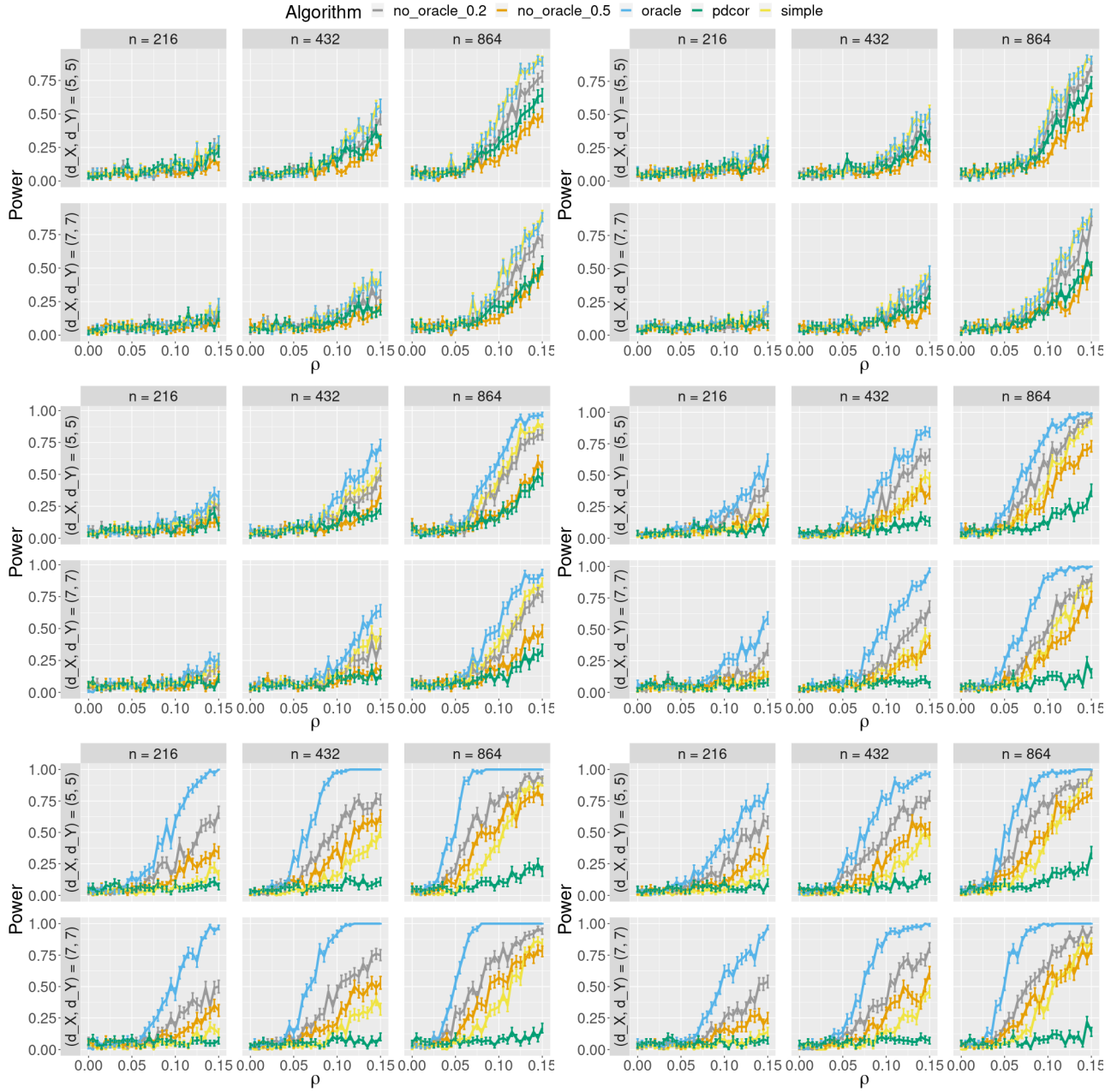


Figure 3: Simulation results for the setting explained in Section B.1. The left 3 and the right 3 columns display the results for $\Sigma^{(1)}$ and $\Sigma^{(2)}$ respectively. The first two rows are for $\tau = 0$, the middle two rows for $\tau = 0.5$, and the last two rows for $\tau = 0.9$. The abbreviation *simple* stands for the approach from Section 3.1, *oracle* for the approach from Section 3.2, *no_oracle_0.2* and *no_oracle_0.5* for the sample split approaches from Section 3.3 with 20% respectively 50% of the sample used for the first part of the algorithm, and *pdcor* for the partial distance correlation.

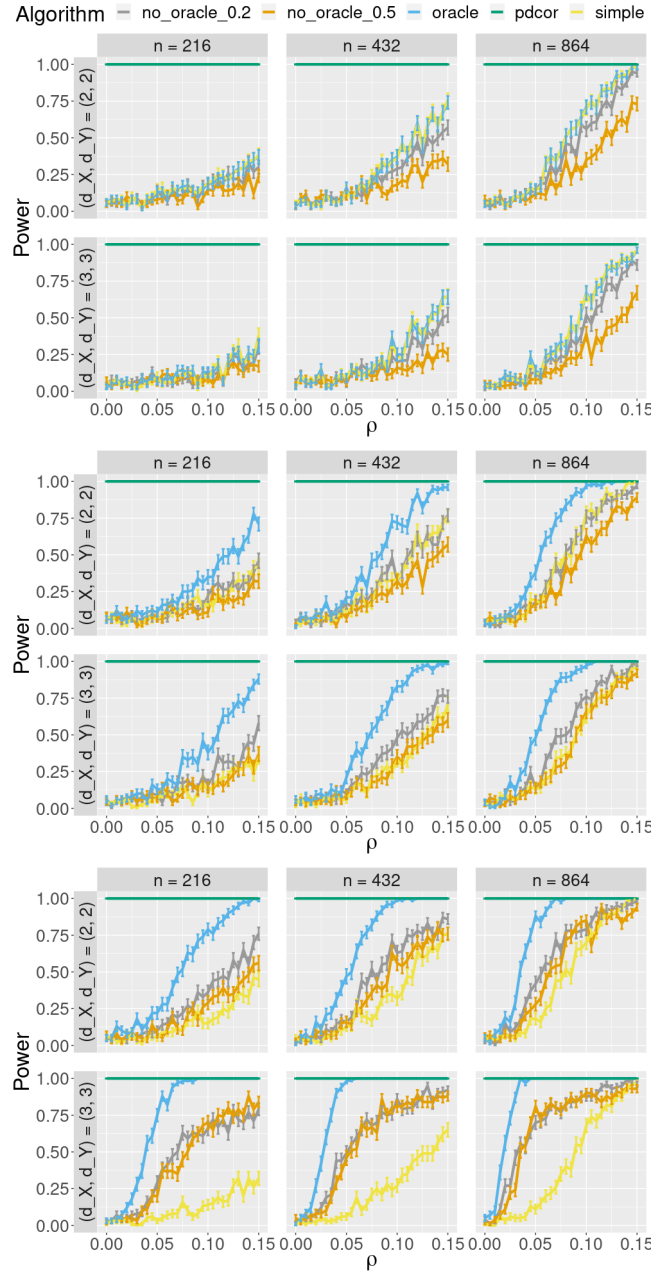


Figure 4: Simulation results for the setting explained in Section B.2. The first two rows are for $\tau = 0$, the middle two rows for $\tau = 0.5$, and the last two rows for $\tau = 0.9$. The abbreviation *simple* stands for the approach from Section 3.1, *oracle* for the approach from Section 3.2, *no_oracle_0.2* and *no_oracle_0.5* for the sample split approaches from Section 3.3 with 20% respectively 50% of the sample used for the first part of the algorithm, and *pdcor* for the partial distance correlation.

(Section 3.3) again performs slightly worse. The partial distance correlation test does not seem to properly control the false positive rate, so we cannot properly compare its power to the other approaches.

C PROOFS

C.1 PROOF OF LEMMA 1

Let $\mathcal{D} \in \mathbb{R}^{n \times (d_X + d_Y + d_Z)}$ be a matrix that has the n rows $(\mathbf{X}^{(1)}, \mathbf{Y}^{(1)}, \mathbf{Z}^{(1)}), \dots, (\mathbf{X}^{(n)}, \mathbf{Y}^{(n)}, \mathbf{Z}^{(n)})$. Let \mathcal{X}^n be the sample space corresponding the observations $(\mathbf{X}^{(1)}, \mathbf{Y}^{(1)}, \mathbf{Z}^{(1)}), \dots, (\mathbf{X}^{(n)}, \mathbf{Y}^{(n)}, \mathbf{Z}^{(n)})$ (where we assume for mathematical rigor that each element of \mathcal{X}^n is a real-valued matrix with n rows and $d_X + d_Y + d_Z$ columns). Let $\psi' : \mathcal{X}^n \rightarrow [0, 1]$ be a (possibly randomized) test for \mathcal{H}'_0 at fixed sample size n and let $\psi : \mathcal{X}^n \rightarrow [0, 1]$ be the induced test for \mathcal{H}_0 . Here 1 corresponds to certain rejection and 0 to certain not-rejection. Denote the set of all possible distributions $\tilde{\mathcal{P}}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}$ for $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ such that \mathcal{H}_0 is true by \mathcal{P}_0 . Similarly, we write \mathcal{P}'_0 for the set of all $\tilde{\mathcal{P}}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}$ such that \mathcal{H}'_0 is true. Moreover, we write $\tilde{\mathcal{P}}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}$ to denote the product measure induced by $\tilde{\mathcal{P}}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}$ for the sample \mathcal{D} , and we write $\mathbb{E}_{\tilde{\mathcal{P}}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}}$ to denote the expectation which is determined by $\tilde{\mathcal{P}}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}$. The size of the test ψ for \mathcal{H}_0 for fixed sample size n is $\sup_{\mathcal{P}_0} \mathbb{E}_{\tilde{\mathcal{P}}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}}[\psi(\mathcal{D})]$.

Now, assume that ψ' has **valid level** for \mathcal{H}'_0 at sample size n , i.e.,

$$\sup_{\mathcal{P}'_0} \mathbb{E}_{\tilde{\mathcal{P}}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}}[\psi'(\mathcal{D})] \leq \alpha.$$

By definition, \mathcal{H}_0 is rejected if and only if \mathcal{H}'_0 had been rejected, therefore,

$$\sup_{\mathcal{P}_0} \mathbb{E}_{\tilde{\mathcal{P}}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}}[\psi(\mathcal{D})] = \sup_{\mathcal{P}_0} \mathbb{E}_{\tilde{\mathcal{P}}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}}[\psi'(\mathcal{D})].$$

Moreover, we note that $\mathcal{P}_0 \subseteq \mathcal{P}'_0$ (by the discussion in Section 2.2) and thus,

$$\sup_{\mathcal{P}_0} \mathbb{E}_{\tilde{\mathcal{P}}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}}[\psi'(\mathcal{D})] \leq \sup_{\mathcal{P}'_0} \mathbb{E}_{\tilde{\mathcal{P}}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}}[\psi'(\mathcal{D})] \leq \alpha,$$

where the last inequality follows from the assumption on the size of ψ' . Hence, ψ has valid level for \mathcal{H}_0 at sample size n .

Now assume that ψ' has **pointwise asymptotic level** for \mathcal{H}'_0 , i.e.,

$$\sup_{\mathcal{P}'_0} \limsup_{n \rightarrow \infty} \mathbb{E}_{\tilde{\mathcal{P}}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}}[\psi'(\mathcal{D})] \leq \alpha.$$

Now, by arguing similarly as before, for all $\tilde{\mathcal{P}}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}} \in \mathcal{P}_0$ we have

$$\limsup_{n \rightarrow \infty} \mathbb{E}_{\tilde{\mathcal{P}}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}}[\psi(\mathcal{D})] = \limsup_{n \rightarrow \infty} \mathbb{E}_{\tilde{\mathcal{P}}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}}[\psi'(\mathcal{D})]$$

and hence

$$\sup_{\mathcal{P}_0} \limsup_{n \rightarrow \infty} \mathbb{E}_{\tilde{\mathcal{P}}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}}[\psi(\mathcal{D})] = \sup_{\mathcal{P}_0} \limsup_{n \rightarrow \infty} \mathbb{E}_{\tilde{\mathcal{P}}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}}[\psi'(\mathcal{D})].$$

Again, $\mathcal{P}_0 \subseteq \mathcal{P}'_0$ and thus,

$$\sup_{\mathcal{P}_0} \limsup_{n \rightarrow \infty} \mathbb{E}_{\tilde{\mathcal{P}}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}}[\psi'(\mathcal{D})] \leq \sup_{\mathcal{P}'_0} \limsup_{n \rightarrow \infty} \mathbb{E}_{\tilde{\mathcal{P}}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}}[\psi'(\mathcal{D})] \leq \alpha.$$

Hence, ψ has pointwise asymptotic level for \mathcal{H}_0 .

Lastly, assume that ψ' has **uniform asymptotic level** for \mathcal{H}'_0 , i.e.,

$$\limsup_{n \rightarrow \infty} \sup_{\mathcal{P}'_0} \mathbb{E}_{\tilde{\mathcal{P}}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}}[\psi'(\mathcal{D})] \leq \alpha.$$

By arguing as before, we have for all $n \in \mathbb{N}$ that

$$\sup_{\mathcal{P}_0} \mathbb{E}_{\tilde{\mathcal{P}}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}}[\psi(\mathcal{D})] = \sup_{\mathcal{P}_0} \mathbb{E}_{\tilde{\mathcal{P}}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}}[\psi'(\mathcal{D})].$$

Again, $\mathcal{P}_0 \subseteq \mathcal{P}'_0$ and thus,

$$\sup_{\mathcal{P}_0} \mathbb{E}_{\tilde{P}_{\mathbf{X},\mathbf{Y},\mathbf{Z}}^{\mathcal{D}}} [\psi'(\mathcal{D})] \leq \sup_{\mathcal{P}'_0} \mathbb{E}_{\tilde{P}_{\mathbf{X},\mathbf{Y},\mathbf{Z}}^{\mathcal{D}}} [\psi'(\mathcal{D})] \leq \alpha$$

for all $n \in \mathbb{N}$. As the previous equality holds for all $n \in \mathbb{N}$,

$$\limsup_{n \rightarrow \infty} \sup_{\mathcal{P}_0} \mathbb{E}_{\tilde{P}_{\mathbf{X},\mathbf{Y},\mathbf{Z}}^{\mathcal{D}}} [\psi'(\mathcal{D})] \leq \limsup_{n \rightarrow \infty} \sup_{\mathcal{P}'_0} \mathbb{E}_{\tilde{P}_{\mathbf{X},\mathbf{Y},\mathbf{Z}}^{\mathcal{D}}} [\psi'(\mathcal{D})] \leq \alpha.$$

Hence, ψ has uniform asymptotic level for \mathcal{H}_0 .

C.2 PROOF OF LEMMA 2

The proof is very similar to the proof of Lemma 1, we reuse most of its notation. Let \mathcal{P}''_0 denote the set of all distributions $\tilde{P}_{\mathbf{X},\mathbf{Y},\mathbf{Z}}$ such that $X_i \perp\!\!\!\perp Y_j \mid (\mathbf{Z}, \mathbf{S}_{ij})$ for all $i \in \{1, \dots, d_X\}$ and $j \in \{1, \dots, d_Y\}$. As shown in the proof sketch in the main paper, $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$ implies $X_i \perp\!\!\!\perp Y_j \mid (\mathbf{Z}, \mathbf{S}_{ij})$ for all $i \in \{1, \dots, d_X\}$ and $j \in \{1, \dots, d_Y\}$, and hence, $\mathcal{P}_0 \subseteq \mathcal{P}''_0$. Because of this fact, the proof of Lemma 2 is the same as the proof of Lemma 1 where \mathcal{P}'_0 is replaced by \mathcal{P}''_0 .

C.3 PROOF OF LEMMA 3

We reuse some notation from the proof of Lemma 1. To emphasize that the test for step 2 depends on the chosen Q_i 's and Q'_j 's, we write $\psi''_{\{Q_i\},\{Q'_j\}}$. Here, $\{Q_i\}$ and $\{Q'_j\}$ are shorthand notation for the sets of all Q_i 's and Q'_j 's respectively. As before, we let ψ denote the induced test procedure for \mathcal{H}_0 . Let $\mathcal{P}''_{0,\{Q_i\},\{Q'_j\}}$ denote the set of all distributions $\tilde{P}_{\mathbf{X},\mathbf{Y},\mathbf{Z}}$ that satisfy $X_i \perp\!\!\!\perp Y_j \mid (\mathbf{Z}, \mathbf{S}_{ij})$ for all $i \in \{1, \dots, d_X\}$ and $j \in \{1, \dots, d_Y\}$, where the \mathbf{S}_{ij} 's are constructed using the respective Q_i 's and Q'_j 's.¹ To underline that the estimates in step 1 depend on the particular sample as well, we write $\hat{S}(X_1, \mathcal{D}), \dots, \hat{S}(X_{d_X}, \mathcal{D}), \hat{S}(Y_1, \mathcal{D}), \dots, \hat{S}(Y_{d_Y}, \mathcal{D})$. We write $A^{\mathcal{D}}_{\{Q_i\},\{Q'_j\}}$ to denote the event $\{(Q_1, \dots, Q_{d_X}, Q'_1, \dots, Q'_{d_Y}) = (\hat{S}(X_1, \mathcal{D}), \dots, \hat{S}(X_{d_X}, \mathcal{D}), \hat{S}(Y_1, \mathcal{D}), \dots, \hat{S}(Y_{d_Y}, \mathcal{D}))\}$, i.e., the event that a particular set of Q_i 's and Q'_j 's has been chosen in step 1. The \mathcal{D} in the superscript of the event indicates that the event is with respect to the random matrix \mathcal{D} . We assume that all these events are measurable, which in practice, depends on the underlying estimation procedure, and we assume without loss of generality that all events have positive measure.

Now, assume that for all possible fixed $Q_1, \dots, Q_{d_X} \subseteq \{1, \dots, d_Y\}$ and $Q'_1, \dots, Q'_{d_Y} \subseteq \{1, \dots, d_X\}$ we have a test in step 2 that has **valid level** $\alpha \in (0, 1)$ conditioned on the fact that $Q_1, \dots, Q_{d_X}, Q'_1, \dots, Q'_{d_Y}$ have been selected in step 1. That means, we assume that for all fixed $Q_1, \dots, Q_{d_X} \subseteq \{1, \dots, d_Y\}$ and $Q'_1, \dots, Q'_{d_Y} \subseteq \{1, \dots, d_X\}$ and for all $\tilde{P}_{\mathbf{X},\mathbf{Y},\mathbf{Z}} \in \mathcal{P}''_{0,\{Q_i\},\{Q'_j\}}$ that

$$\mathbb{E}_{\tilde{P}_{\mathbf{X},\mathbf{Y},\mathbf{Z}}^{\mathcal{D}}} [\psi''_{\{Q_i\},\{Q'_j\}}(\mathcal{D}) \mid A^{\mathcal{D}}_{\{Q_i\},\{Q'_j\}}] \leq \alpha. \quad (1)$$

(This notion of conditioning is similar to the one from Fithian et al. [2014]). By the proof sketch in the main paper, $\mathcal{P}_0 \subseteq \mathcal{P}''_{0,\{Q_i\},\{Q'_j\}}$ for all $Q_1, \dots, Q_{d_X} \subseteq \{1, \dots, d_Y\}$ and $Q'_1, \dots, Q'_{d_Y} \subseteq \{1, \dots, d_X\}$ (and not just for all $Q_i \subseteq S(X_i)$ and $Q'_j \subseteq S(Y_j)$) and hence, inequality (1) holds for all $\tilde{P}_{\mathbf{X},\mathbf{Y},\mathbf{Z}} \in \mathcal{P}_0$ as well. Thus, by applying the law of total expectation,

¹Actually, $\psi''_{\{Q_i\},\{Q'_j\}}$ and $\mathcal{P}''_{0,\{Q_i\},\{Q'_j\}}$ only depend on the Q_i 's and Q'_j 's that are chosen for some \mathbf{S}_{ij} . For better clarification, we do not make this distinction in our notation.

we have for all $\tilde{P}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}} \in \mathcal{P}_0$ that

$$\begin{aligned}
& \mathbb{E}_{\tilde{P}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}}[\psi(\mathcal{D})] \\
&= \sum_{\substack{Q_1, \dots, Q_{d_X} \subseteq \{1, \dots, d_Y\} \\ Q'_1, \dots, Q'_{d_Y} \subseteq \{1, \dots, d_X\}}} \mathbb{E}_{\tilde{P}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}}[\psi(\mathcal{D}) \mid A_{\{Q_i\}, \{Q'_j\}}^{\mathcal{D}}] \cdot \tilde{P}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}(A_{\{Q_i\}, \{Q'_j\}}^{\mathcal{D}}) \\
&= \sum_{\substack{Q_1, \dots, Q_{d_X} \subseteq \{1, \dots, d_Y\} \\ Q'_1, \dots, Q'_{d_Y} \subseteq \{1, \dots, d_X\}}} \mathbb{E}_{\tilde{P}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}}[\psi''_{\{Q_i\}, \{Q'_j\}}(\mathcal{D}) \mid A_{\{Q_i\}, \{Q'_j\}}^{\mathcal{D}}] \cdot \tilde{P}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}(A_{\{Q_i\}, \{Q'_j\}}^{\mathcal{D}}) \\
&\leq \sum_{\substack{Q_1, \dots, Q_{d_X} \subseteq \{1, \dots, d_Y\} \\ Q'_1, \dots, Q'_{d_Y} \subseteq \{1, \dots, d_X\}}} \alpha \cdot \tilde{P}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}(A_{\{Q_i\}, \{Q'_j\}}^{\mathcal{D}}) \\
&= \alpha \sum_{\substack{Q_1, \dots, Q_{d_X} \subseteq \{1, \dots, d_Y\} \\ Q'_1, \dots, Q'_{d_Y} \subseteq \{1, \dots, d_X\}}} \tilde{P}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}(A_{\{Q_i\}, \{Q'_j\}}^{\mathcal{D}}) \\
&= \alpha,
\end{aligned}$$

where we used twice that the events $A_{\{Q_i\}, \{Q'_j\}}^{\mathcal{D}}$ partition the underlying sample space \mathcal{X}^n . Hence,

$$\sup_{\mathcal{P}_0} \mathbb{E}_{\tilde{P}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}}[\psi(\mathcal{D})] \leq \alpha.$$

For the **sample splitting part** of the Lemma, we write \mathcal{D}_1 to denote a matrix that is constructed by an arbitrary but fixed subset of rows of the matrix \mathcal{D} , and we define \mathcal{D}_2 to contain exactly all the other rows that do not make up \mathcal{D}_1 . Without loss of generality, assume that \mathcal{D}_1 is made up of the first n_1 rows of \mathcal{D} and \mathcal{D}_2 of the remaining $n_2 := n - n_1$ rows. Now, suppose that the estimates $\hat{S}(X_1, \mathcal{D}), \dots, \hat{S}(X_{d_X}, \mathcal{D}), \hat{S}(Y_1, \mathcal{D}), \dots, \hat{S}(Y_{d_Y}, \mathcal{D})$ are calculated using only the first part of the sample \mathcal{D}_1 and suppose that the test $\psi''_{\{Q_i\}, \{Q'_j\}}(\mathcal{D})$ is calculated based on \mathcal{D}_2 only.

Under the assumption that all observations are mutually independent (i.e, under the classical iid assumption), we get that $\psi''_{\{Q_i\}, \{Q'_j\}}(\mathcal{D})$ is independent of $(\hat{S}(X_1, \mathcal{D}), \dots, \hat{S}(X_{d_X}, \mathcal{D}), \hat{S}(Y_1, \mathcal{D}), \dots, \hat{S}(Y_{d_Y}, \mathcal{D}))$. Because $\psi''_{\{Q_i\}, \{Q'_j\}}(\mathcal{D})$ only depends on \mathcal{D}_2 , we can define a test $\phi''_{\{Q_i\}, \{Q'_j\}} : \mathcal{X}^{|\mathcal{D}_2|} \rightarrow [0, 1]$ that is equal to $\psi''_{\{Q_i\}, \{Q'_j\}}(\mathcal{D})$ in the sense that $\psi''_{\{Q_i\}, \{Q'_j\}}(\mathcal{D}) = \psi''_{\{Q_i\}, \{Q'_j\}}(\mathcal{D}_1, \mathcal{D}_2) = \phi''_{\{Q_i\}, \{Q'_j\}}(\mathcal{D}_2)$ (this definition is just a technical detail, loosely speaking, we could directly write $\psi''_{\{Q_i\}, \{Q'_j\}}(\mathcal{D}_2)$ to indicate that the test depends on the dataset \mathcal{D}_2 only, but we need to make sure that the underlying domain is correct). We now obtain that

$$\begin{aligned}
& \mathbb{E}_{\tilde{P}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}}[\psi''_{\{Q_i\}, \{Q'_j\}}(\mathcal{D}) \mid A_{\{Q_i\}, \{Q'_j\}}^{\mathcal{D}}] \\
&= \mathbb{E}_{\tilde{P}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}}[\psi''_{\{Q_i\}, \{Q'_j\}}(\mathcal{D})] \\
&= \mathbb{E}_{\tilde{P}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}}[\phi''_{\{Q_i\}, \{Q'_j\}}(\mathcal{D}_2)] \\
&= \mathbb{E}_{\tilde{P}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}_2}}[\phi''_{\{Q_i\}, \{Q'_j\}}(\mathcal{D}_2)]. \tag{2}
\end{aligned}$$

The term $\mathbb{E}_{\tilde{P}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}_2}}[\phi''_{\{Q_i\}, \{Q'_j\}}(\mathcal{D}_2)]$ in (2) is the usual unconditional size of the test based on the second part of the dataset. Thus, if for all fixed $Q_1, \dots, Q_{d_X} \subseteq \{1, \dots, d_Y\}$ and $Q'_1, \dots, Q'_{d_Y} \subseteq \{1, \dots, d_X\}$ it holds that

$$\sup_{\mathcal{P}''_{0, \{Q_i\}, \{Q'_j\}}} \mathbb{E}_{\tilde{P}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}_2}}[\phi''_{\{Q_i\}, \{Q'_j\}}(\mathcal{D}_2)] \leq \alpha,$$

we obtain that

$$\sup_{\mathcal{P}_0} \mathbb{E}_{\tilde{P}_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}}[\psi(\mathcal{D})] \leq \alpha.$$

C.4 ANALOGOUS RESULT AS IN PROPOSITION 1 FOR PARTIAL CORRELATIONS

In this section, we prove an analogous result as in Proposition 1 for partial correlations.

Proposition 4. For any set of indices $Q_i \subseteq S(X_i)$,

$$|\rho_{X_i Y_j | \mathbf{Z}, \mathbf{Y}_{Q_i \setminus \{j\}}} | \geq |\rho_{X_i Y_j | \mathbf{Z}} |.$$

Similarly, for any set of indices $Q'_j \subseteq S(Y_j)$,

$$|\rho_{X_i Y_j | \mathbf{Z}, \mathbf{X}_{Q'_j \setminus \{i\}}} | \geq |\rho_{X_i Y_j | \mathbf{Z}} |$$

Proof. We only prove the statement for any arbitrary but fixed $Q_i \subseteq S(X_i)$, the proof for any arbitrary but fixed $Q'_j \subseteq S(Y_j)$ is analogous.

Write $S(X_i) \setminus \{j\} = \{j_1, \dots, j_m\}$, where m is a natural number such that $1 \leq m \leq d_Y - 1$. Without loss of generality (as we can relabel the elements j_1, \dots, j_m arbitrarily), we prove the statement (and a slightly stronger statement for later use) for all sets $\{j_1, \dots, j_k\} \subseteq S(X_i)$ by induction over $k \in \{1, \dots, m\}$.²

Induction start ($k = 1$): First of all, note that $X_i \perp\!\!\!\perp Y_{j_l} | \mathbf{Z}$ for all $j_l \in S(X_i) \setminus \{j\}$, hence, $\rho_{X_i Y_{j_l} | \mathbf{Z}} = 0$ for all $j_l \in S(X_i)$. Therefore,

$$\begin{aligned} \rho_{X_i Y_j | \mathbf{Z}, Y_{j_1}} &= \frac{\rho_{X_i Y_j | \mathbf{Z}} - \overbrace{\rho_{X_i Y_{j_1} | \mathbf{Z}} \rho_{Y_j Y_{j_1} | \mathbf{Z}}}^{=0}}{\underbrace{\sqrt{1 - \rho_{X_i Y_{j_1} | \mathbf{Z}}^2}}_{=1} \sqrt{1 - \rho_{Y_j Y_{j_1} | \mathbf{Z}}^2}} \\ &= \frac{\rho_{X_i Y_j | \mathbf{Z}}}{\sqrt{1 - \rho_{Y_j Y_{j_1} | \mathbf{Z}}^2}} \end{aligned}$$

and hence,

$$|\rho_{X_i Y_j | \mathbf{Z}, Y_{j_1}} | \geq |\rho_{X_i Y_j | \mathbf{Z}} |$$

Induction hypothesis: Let $k \leq m$ be an arbitrary but fixed natural number. Assume that

$$\rho_{X_i Y_{j_l} | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_{k-1}}} = 0 \quad (3)$$

for all $l \in \{k, \dots, m\}$. Here, the notation $Y_{j_1}, \dots, Y_{j_{k-1}}$ means the empty set if $k = 1$.

Furthermore, assume that

$$\rho_{X_i Y_j | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_k}} = \frac{\rho_{X_i Y_j | \mathbf{Z}}}{\sqrt{1 - \rho_{Y_j Y_{j_1} | \mathbf{Z}}^2} \sqrt{1 - \rho_{Y_j Y_{j_2} | \mathbf{Z}, Y_{j_1}}^2} \cdots \sqrt{1 - \rho_{Y_j Y_{j_k} | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_{k-1}}}^2}} \quad (4)$$

and hence that

$$|\rho_{X_i Y_j | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_k}} | \geq |\rho_{X_i Y_j | \mathbf{Z}} |.$$

Induction step ($k \rightarrow k + 1$): If $k = m$, then we are done, so assume that $k < m$. Let $l \in \{k + 1, \dots, m\}$ be arbitrary but fixed. Then,

$$\begin{aligned} \rho_{X_i Y_{j_l} | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_k}} &= \frac{\overbrace{\rho_{X_i Y_{j_l} | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_{k-1}}} }^{=0, \text{ equation (3)}} - \overbrace{\rho_{X_i Y_{j_k} | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_{k-1}}} \rho_{Y_{j_k} Y_{j_l} | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_{k-1}}} }^{=0, \text{ equation (3)}}}{\underbrace{\sqrt{1 - \rho_{X_i Y_{j_k} | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_{k-1}}}^2}}_{=1, \text{ equation (3)}} \sqrt{1 - \rho_{Y_{j_k} Y_{j_l} | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_{k-1}}}^2}} \\ &= \frac{0 - 0 \cdot \rho_{Y_{j_k} Y_{j_l} | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_{k-1}}}}{1 \cdot \sqrt{1 - \rho_{Y_{j_k} Y_{j_l} | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_{k-1}}}^2}} \\ &= 0. \end{aligned} \quad (5)$$

²Usually, a proof by induction is over the entire natural numbers. To be formally correct, we can also say that we inductively proof a statement that equals the original statement for $k \in \{1, \dots, m\}$ and is always correct for $k > m$.

With this, we obtain that

$$\begin{aligned}
\rho_{X_i Y_j | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_{k+1}}} &= \frac{\overbrace{\rho_{X_i Y_j | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_k}} - \rho_{X_i Y_{j_{k+1}} | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_k}} \rho_{Y_j Y_{j_{k+1}} | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_k}}}^{=0, \text{ equation (5)}}}{\underbrace{\sqrt{1 - \rho_{X_i Y_{j_{k+1}} | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_k}}^2}}_{=1, \text{ equation (5)}} \sqrt{1 - \rho_{Y_j Y_{j_{k+1}} | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_k}}^2}} \\
&= \frac{\rho_{X_i Y_j | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_k}} - 0 \cdot \rho_{Y_j Y_{j_{k+1}} | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_k}}}{1 \cdot \sqrt{1 - \rho_{Y_j Y_{j_{k+1}} | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_k}}^2}} \\
&= \frac{\rho_{X_i Y_j | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_k}}}{\sqrt{1 - \rho_{Y_j Y_{j_{k+1}} | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_k}}^2}} \\
&= \frac{\rho_{X_i Y_j | \mathbf{Z}}}{\sqrt{1 - \rho_{Y_j Y_{j_1} | \mathbf{Z}}^2} \sqrt{1 - \rho_{Y_j Y_{j_2} | \mathbf{Z}, Y_{j_1}}^2} \cdots \sqrt{1 - \rho_{Y_j Y_{j_{k+1}} | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_k}}^2}}
\end{aligned}$$

where the last equality follows from the induction hypothesis, see equation (4). Therefore,

$$|\rho_{X_i Y_j | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_{k+1}}| \geq |\rho_{X_i Y_j | \mathbf{Z}}|.$$

□

C.5 PROOF OF PROPOSITION 2

We start the proof by calculating the power corresponding to the null hypothesis $\rho_{X_i Y_j | \mathbf{Z}} = 0$. One rejects that null hypothesis if $\sqrt{n-3-|\mathbf{Z}|} |z(\hat{\rho}_{X_i Y_j | \mathbf{Z}})| > q$, where $q := \Phi^{-1}(1 - \alpha/2)$ is the $1 - \alpha/2$ -quantile of a standard normal distribution. Assume that the true $\rho_{X_i Y_j | \mathbf{Z}}$ is fixed and not equal to zero. Note that this fact is equivalent to $z(\rho_{X_i Y_j | \mathbf{Z}}) \neq 0$. The statistical power is then

$$\begin{aligned}
&P_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}(\sqrt{n-3-|\mathbf{Z}|} |z(\hat{\rho}_{X_i Y_j | \mathbf{Z}})| > q) \\
&= P_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}(\sqrt{n-3-|\mathbf{Z}|} |z(\hat{\rho}_{X_i Y_j | \mathbf{Z}})| > q \text{ or } \sqrt{n-3-|\mathbf{Z}|} |z(\hat{\rho}_{X_i Y_j | \mathbf{Z}})| < -q) \\
&= P_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}(\sqrt{n-3-|\mathbf{Z}|} |z(\hat{\rho}_{X_i Y_j | \mathbf{Z}})| > q) + P_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}(\sqrt{n-3-|\mathbf{Z}|} |z(\hat{\rho}_{X_i Y_j | \mathbf{Z}})| < -q) \\
&= P_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}(\sqrt{n-3-|\mathbf{Z}|} |(z(\hat{\rho}_{X_i Y_j | \mathbf{Z}}) - z(\rho_{X_i Y_j | \mathbf{Z}}))| > q - \sqrt{n-3-|\mathbf{Z}|} |z(\rho_{X_i Y_j | \mathbf{Z}})|) \\
&\quad + P_{\mathbf{X}, \mathbf{Y}, \mathbf{Z}}^{\mathcal{D}}(\sqrt{n-3-|\mathbf{Z}|} |(z(\hat{\rho}_{X_i Y_j | \mathbf{Z}}) - z(\rho_{X_i Y_j | \mathbf{Z}}))| < -q - \sqrt{n-3-|\mathbf{Z}|} |z(\rho_{X_i Y_j | \mathbf{Z}})|) \\
&=: (1).
\end{aligned}$$

Using that the left-hand-side in the probabilities is (approximately) standard normally distributed, we obtain that

$$(1) = P(W > q - \gamma_1) + P(W < -q - \gamma_1),$$

where W is a standard normal random variable, P is the underlying probability measure, and $\gamma_1 = \sqrt{n-3-|\mathbf{Z}|} |z(\rho_{X_i Y_j | \mathbf{Z}})|$.

By the same argument, we can calculate the power corresponding to the null hypothesis $\rho_{X_i Y_j | \mathbf{Z}, \mathbf{Y}_{Q_i \setminus \{j\}}} = 0$, (note that this null hypothesis is true if and only if $\rho_{X_i Y_j | \mathbf{Z}} = 0$; for that, see the proof of Proposition 4). These calculations yield that the power (approximately) equals

$$(2) := P(W > q - \gamma_2) + P(W < -q - \gamma_2),$$

where $\gamma_2 = \sqrt{n_2-3-|\mathbf{Z}| - |Q_i \setminus \{j\}|} |z(\rho_{X_i Y_j | \mathbf{Z}, \mathbf{Y}_{Q_i \setminus \{j\}})|$. The power difference $\Delta\beta$ between (2) and (1) then is

$$\Delta\beta = P(W > q - \gamma_2) + P(W < -q - \gamma_2) - P(W > q - \gamma_1) - P(W < -q - \gamma_1).$$

We now use our assumption that

$$I(Y_j; \mathbf{Y}_{Q_i \setminus \{j\}} | \mathbf{Z}) \geq \log \left(\frac{z^{-1} \left(\sqrt{\frac{n-3-|\mathbf{Z}|}{n_2-3-|\mathbf{Z}|-|Q_i \setminus \{j\}|}} z(\rho_{X_i Y_j | \mathbf{Z}}) \right)}{\rho_{X_i Y_j | \mathbf{Z}}} \right). \quad (6)$$

We start by assuming that $\rho_{X_i Y_j | \mathbf{Z}} > 0$ (which, see the proof of Proposition 4, is equivalent to $\rho_{X_i Y_j | \mathbf{Z}, \mathbf{Y}_{Q_i \setminus \{j\}}} > 0$). With that assumption, we can rearrange (6) and obtain that

$$e^{I(Y_j; \mathbf{Y}_{Q_i \setminus \{j\}} | \mathbf{Z})} \rho_{X_i Y_j | \mathbf{Z}} \geq z^{-1} \left(\sqrt{\frac{n-3-|\mathbf{Z}|}{n_2-3-|\mathbf{Z}|-|Q_i \setminus \{j\}|}} z(\rho_{X_i Y_j | \mathbf{Z}}) \right).$$

Writing $Q_i \setminus \{j\} = \{j_1, \dots, j_m\}$ and using the chain rule for conditional mutual information, we obtain that

$$e^{I(Y_j; Y_{j_1} | \mathbf{Z}) + I(Y_j; Y_{j_2} | \mathbf{Z}, Y_{j_1}) + \dots + I(Y_j; Y_{j_m} | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_{m-1}})} \rho_{X_i Y_j | \mathbf{Z}} \geq z^{-1} \left(\sqrt{\frac{n-3-|\mathbf{Z}|}{n_2-3-|\mathbf{Z}|-|Q_i \setminus \{j\}|}} z(\rho_{X_i Y_j | \mathbf{Z}}) \right).$$

Further rearrangements yield

$$e^{I(Y_j; Y_{j_1} | \mathbf{Z})} \cdot e^{I(Y_j; Y_{j_2} | \mathbf{Z}, Y_{j_1})} \dots e^{I(Y_j; Y_{j_m} | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_{m-1}})} \rho_{X_i Y_j | \mathbf{Z}} \geq z^{-1} \left(\sqrt{\frac{n-3-|\mathbf{Z}|}{n_2-3-|\mathbf{Z}|-|Q_i \setminus \{j\}|}} z(\rho_{X_i Y_j | \mathbf{Z}}) \right).$$

Recall that for multivariate normal distributions, $e^{I(Y_j; Y_{j_1} | \mathbf{Z})} = 1 / \sqrt{1 - \rho_{Y_j Y_{j_1} | \mathbf{Z}}^2}$ (analogously for the other terms), which then yields that

$$\frac{\rho_{X_i Y_j | \mathbf{Z}}}{\sqrt{1 - \rho_{Y_j Y_{j_1} | \mathbf{Z}}^2} \cdot \sqrt{1 - \rho_{Y_j Y_{j_2} | \mathbf{Z}, Y_{j_1}}^2} \cdots \sqrt{1 - \rho_{Y_j Y_{j_m} | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_{m-1}}}^2}} \geq z^{-1} \left(\sqrt{\frac{n-3-|\mathbf{Z}|}{n_2-3-|\mathbf{Z}|-|Q_i \setminus \{j\}|}} z(\rho_{X_i Y_j | \mathbf{Z}}) \right).$$

Because $X_i \perp\!\!\!\perp Y_{j_l} | \mathbf{Z}$ for all $l \in \{1, \dots, m\}$, we can (by looking at the proof of Proposition 4) simplify the left-hand side of the previous inequality to

$$\frac{\rho_{X_i Y_j | \mathbf{Z}}}{\sqrt{1 - \rho_{Y_j Y_{j_1} | \mathbf{Z}}^2} \cdot \sqrt{1 - \rho_{Y_j Y_{j_2} | \mathbf{Z}, Y_{j_1}}^2} \cdots \sqrt{1 - \rho_{Y_j Y_{j_m} | \mathbf{Z}, Y_{j_1}, \dots, Y_{j_{m-1}}}^2}} = \rho_{X_i Y_j | \mathbf{Z}, \mathbf{Y}_{Q_i \setminus \{j\}}}.$$

Thus,

$$\rho_{X_i Y_j | \mathbf{Z}, \mathbf{Y}_{Q_i \setminus \{j\}}} \geq z^{-1} \left(\sqrt{\frac{n-3-|\mathbf{Z}|}{n_2-3-|\mathbf{Z}|-|Q_i \setminus \{j\}|}} z(\rho_{X_i Y_j | \mathbf{Z}}) \right).$$

Further rearrangement then yields that

$$\sqrt{n_2-3-|\mathbf{Z}|-|Q_i \setminus \{j\}|} z(\rho_{X_i Y_j | \mathbf{Z}, \mathbf{Y}_{Q_i \setminus \{j\}}}) \geq \sqrt{n-3-|\mathbf{Z}|} z(\rho_{X_i Y_j | \mathbf{Z}})$$

which implies that

$$0 < \gamma_1 \leq \gamma_2.$$

Thus,

$$\begin{aligned} \Delta\beta &= P(q - \gamma_2 < W < q - \gamma_1) - P(-q - \gamma_2 < W < -q - \gamma_1) \\ &= P(\gamma_1 - q < W < \gamma_2 - q) - P(\gamma_1 + q < W < \gamma_2 + q). \end{aligned}$$

To see that $\Delta\beta$ is positive, we make a case distinction. First of all note that $\gamma_2 + q$ and $\gamma_1 + q$ are always positive because we assumed that $\rho_{X_i Y_j | \mathbf{Z}}$ is positive, and hence (see the proof of Proposition 4), $\rho_{X_i Y_j | \mathbf{Z}, \mathbf{Y}_{Q_i \setminus \{j\}}}$ is positive.

If $\gamma_1 - q \geq 0$, both probabilities are integrals of the standard normal density over intervals which are a subset of the nonnegative part of the real line. As the standard normal density is strictly decreasing on the nonnegative real line and

both the interval $(\gamma_1 - q, \gamma_2 - q)$ and $(\gamma_1 + q, \gamma_2 + q)$ have the same length, the first probability is larger than the second probability.

If $\gamma_1 - q < 0$ and $\gamma_2 - q \geq 0$, the interval $(\gamma_1 + q, \gamma_2 + q)$ is further away from zero than the interval $(\gamma_1 - q, \gamma_2 - q)$; and hence, the standard normal density takes smaller values on $(\gamma_1 + q, \gamma_2 + q)$ than on $(\gamma_1 - q, \gamma_2 - q)$. As again both integrals have the same length, the first probability is larger than the second probability.

The last case is $\gamma_2 - q < 0$ (and hence, $\gamma_1 - q < 0$). Note that $\gamma_2 - q$ is closer to zero than $\gamma_1 + q$, because

$$\begin{aligned} |\gamma_2 - q| &< |\gamma_1 + q| \\ \iff q - \gamma_2 &< \gamma_1 + q \\ \iff \gamma_1 + \gamma_2 &> 0, \end{aligned}$$

where the last line is always true by definition of γ_1 and γ_2 . Hence, the interval $(\gamma_1 - q, \gamma_2 - q)$ is again closer to zero than the interval $(\gamma_1 + q, \gamma_2 + q)$, and hence, the first probability is again greater than the second probability.

Thus, in all cases, we have $\Delta\beta \geq 0$.

If $\rho_{X_i Y_j | \mathbf{Z}}$ is negative, then one can proceed analogously and obtain that $\gamma_2 \leq \gamma_1 < 0$. It then follows that

$$\Delta\beta = P(-q - \gamma_1 < W < -q - \gamma_2) - P(q - \gamma_1 < W < q - \gamma_2).$$

By a similar case distinction, one then obtains that $\Delta\beta \geq 0$. Thus, we have proved the proposition.

C.6 PROOF OF PROPOSITION 3

Let both tests corresponding to the respective null hypotheses $\rho_{X_i Y_j | \mathbf{Z}} = 0$ and $\rho_{X_i Y_j | \mathbf{Z}, \mathbf{Y}_{Q_i \setminus \{j\}}} = 0$ achieve a power of exactly β . From the proof of Proposition 2, we use the formulas for calculating the power. We start by assuming that $\rho_{X_i Y_j | \mathbf{Z}} > 0$ (equivalent to $\rho_{X_i Y_j | \mathbf{Z}, \mathbf{Y}_{Q_i \setminus \{j\}}} > 0$, see proof of Proposition 4) and obtain that for the test corresponding to the null hypothesis $\rho_{X_i Y_j | \mathbf{Z}} = 0$,

$$\begin{aligned} \beta &= P(W > q - \gamma_1) + P(W < -q - \gamma_1) \\ &\leq P(W > q - \gamma_1) + P(W < -q) \\ &= P(W > q - \gamma_1) + \frac{\alpha}{2}. \end{aligned}$$

Rearranging terms yields

$$\gamma_1 \geq q - \Phi^{-1}(1 - \beta + \alpha/2).$$

Note that the term on the right-hand side is positive because we assumed that $\beta \geq \alpha$. Plugging in the definition of γ_1 and further rearrangement now yields that

$$n \geq \left(\frac{q - \Phi^{-1}(1 - \beta + \frac{\alpha}{2})}{z(\rho_{X_i Y_j | \mathbf{Z}})} \right)^2 + 3 + |\mathbf{Z}|. \quad (7)$$

Similarly, for the test corresponding to the null hypothesis $\rho_{X_i Y_j | \mathbf{Z}, \mathbf{Y}_{Q_i \setminus \{j\}}} = 0$,

$$\beta = P(W > q - \gamma_2) + P(W < -q - \gamma_2) \geq P(W > q - \gamma_2).$$

Rearranging these terms yields

$$q - \Phi^{-1}(1 - \beta) \geq \gamma_2 \geq 0.$$

Plugging in the definition of γ_2 and rearranging yields

$$n_2 \leq \left(\frac{q - \Phi^{-1}(1 - \beta)}{z(\rho_{X_i Y_j | \mathbf{Z}, \mathbf{Y}_{Q_i \setminus \{j\}})} \right)^2 + 3 + |\mathbf{Z}| + |Q_i \setminus \{j\}|. \quad (8)$$

One can work similarly for the case $\rho_{X_i Y_j | \mathbf{Z}} < 0$ (equivalent to $\rho_{X_i Y_j | \mathbf{Z}, \mathbf{Y}_{Q_i \setminus \{j\}}} < 0$, see the proof of Proposition 4) and obtain the same lower bound on n and upper bound on n_2 .

Putting together (7) and (8) then yields the lower bound on $n - n_2$ stated in the proposition.

References

- Theodore W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, Hoboken, New Jersey, 3rd edition, 2003.
- William Fithian, Dennis L. Sun, and Jonathan E. Taylor. Optimal inference after model selection. *arXiv*, abs/1410.2597, 2014.
- Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, Cambridge, UK, 2nd edition, 2009.
- Jonas Peters and Rajen D. Shah. *GeneralisedCovarianceMeasure: Test for Conditional Independence Based on the Generalized Covariance Measure (GCM)*, 2022. URL <https://CRAN.R-project.org/package=GeneralisedCovarianceMeasure>. R package version 0.2.0.
- Rajen Dinesh Shah and Jonas Peters. The hardness of conditional independence testing and the generalised covariance measure. *The Annals of Statistics*, 48(3):1514–1538, 2020.
- Peter Spirtes, Clark N. Glymour, Richard Scheines, and David Heckerman. *Causation, prediction, and search*. MIT press, 2nd edition, 2000.