Supplementary Material

Conditional independence testing based on a nearest-neighbor estimator of conditional mutual information

Jakob Runge

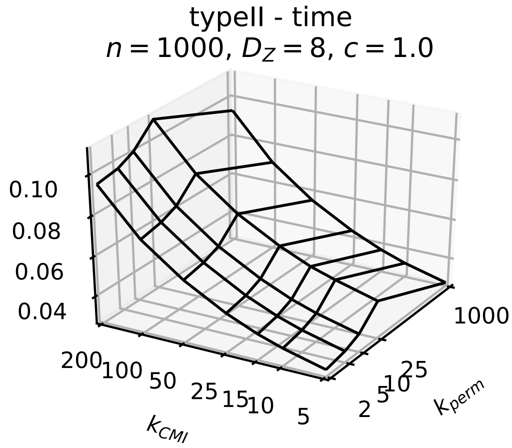

Figure S1: Runtime per estimate [in s] for the same setup as in Fig. 3 but with $n = 1,000$. For $k_{\mathrm{perm}} = n$ a computationally cheaper full permutation scheme was used.

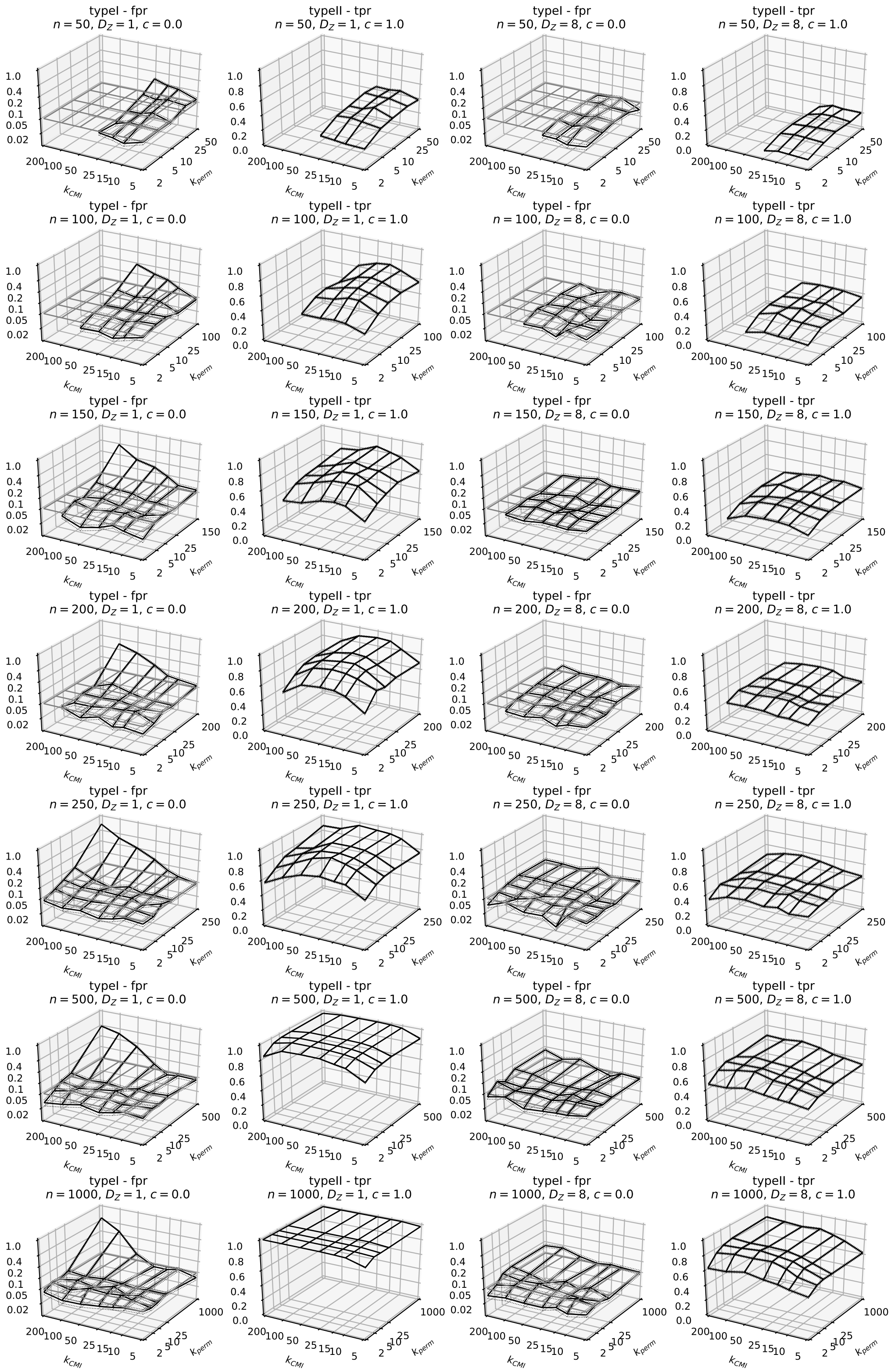

Figure S2: Same as in Fig. 3, but for more sample sizes from $n = 50$ (top) to $n = 1000$ (bottom).

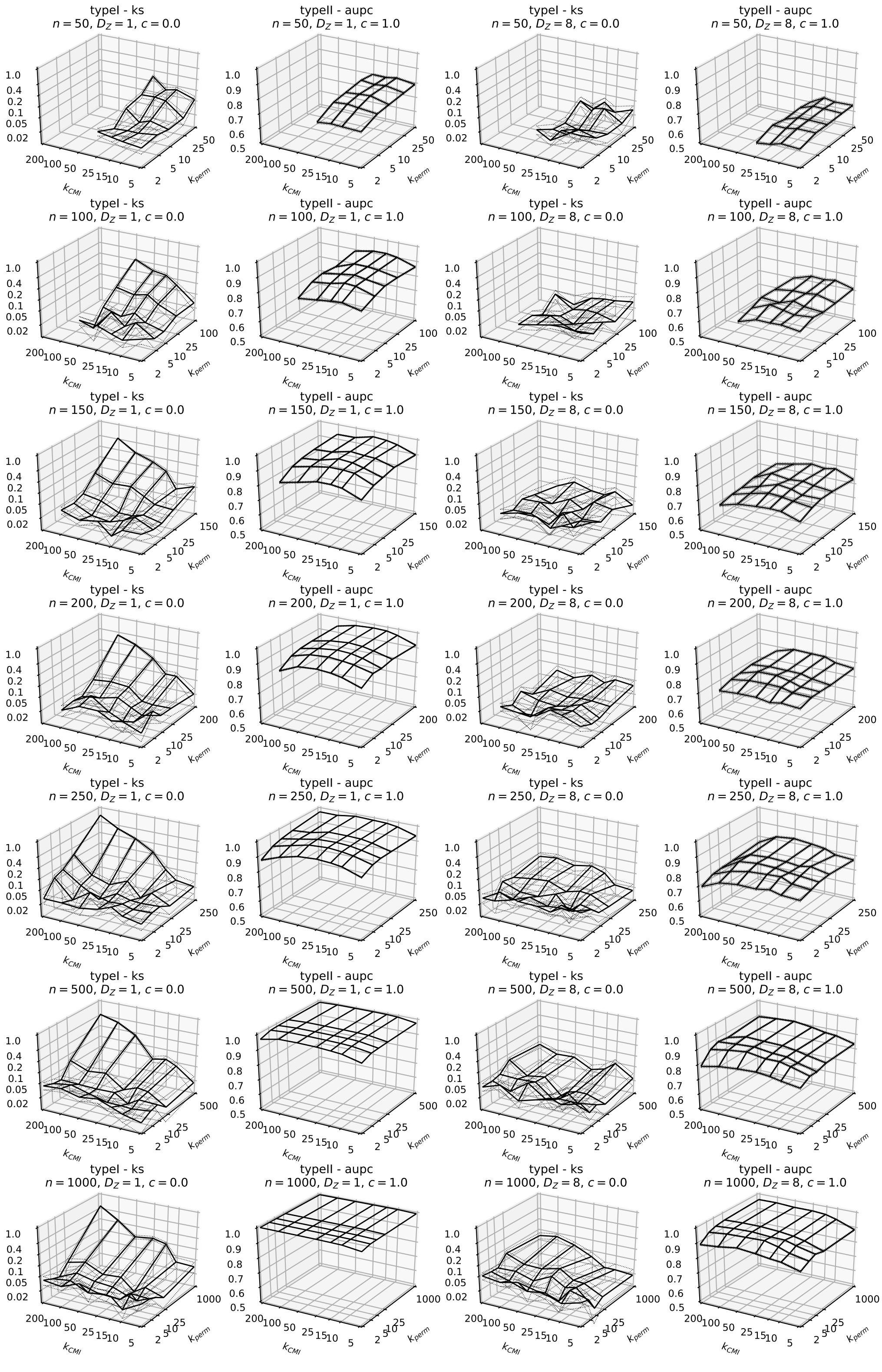

Figure S3: Same as in Fig. 3, but for KS and AUPC metrics and more sample sizes from $n = 50$ (top) to $n = 1000$ (bottom).

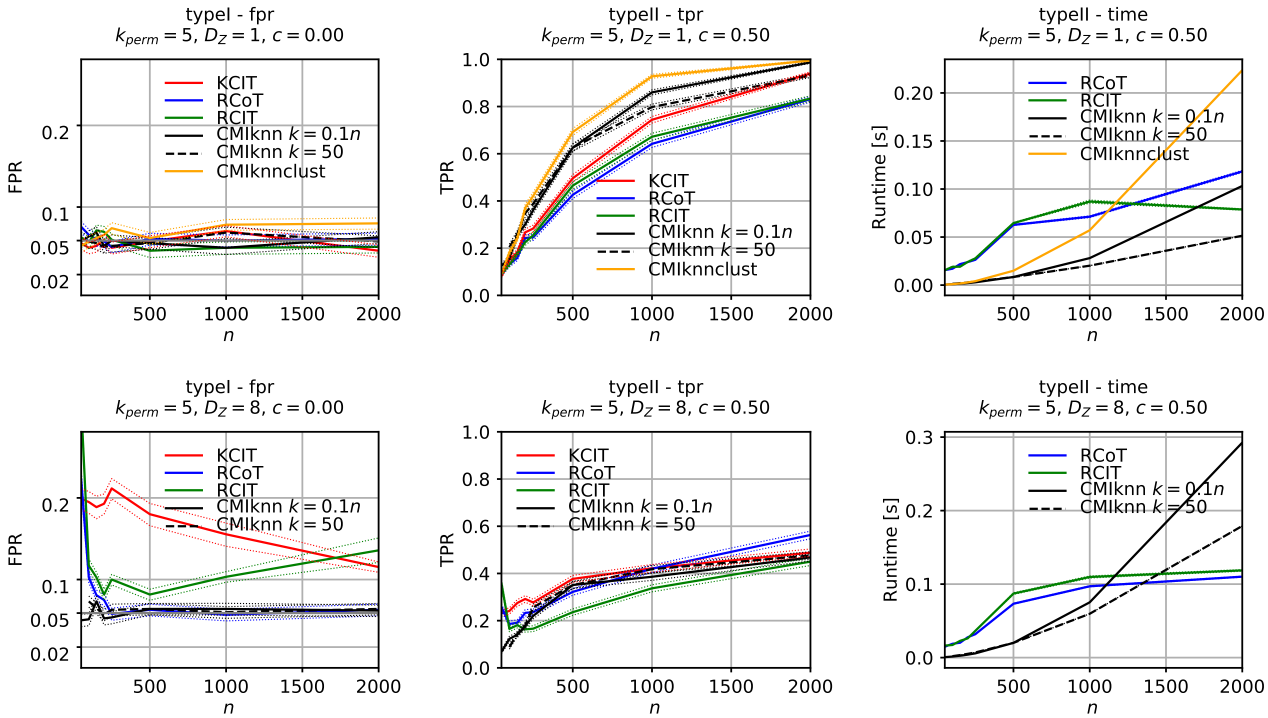

Figure S4: As in Fig. 4 but for false positive rates (FPR) and true positive rates (TPR) at an $\alpha = 0.05$ significance level.

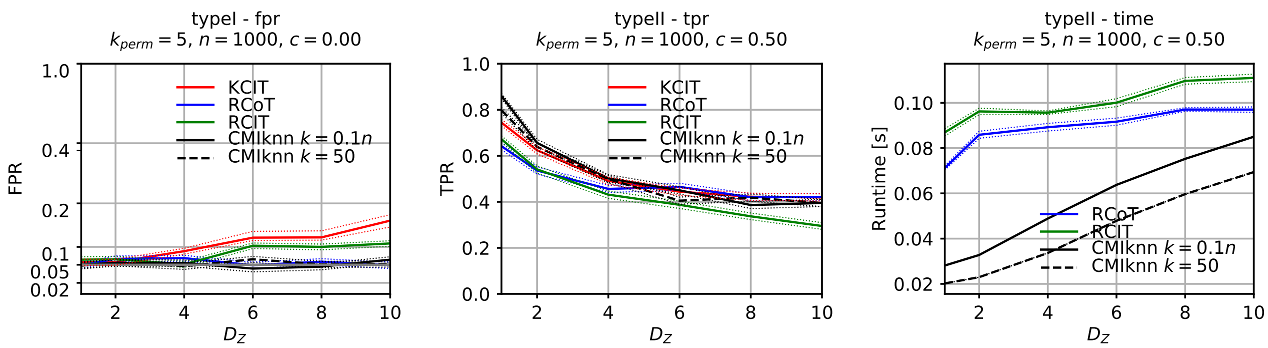

Figure S5: As in Fig. 5 but for false positive rates (FPR) and true positive rates (TPR) at an $\alpha = 0.05$ significance level.

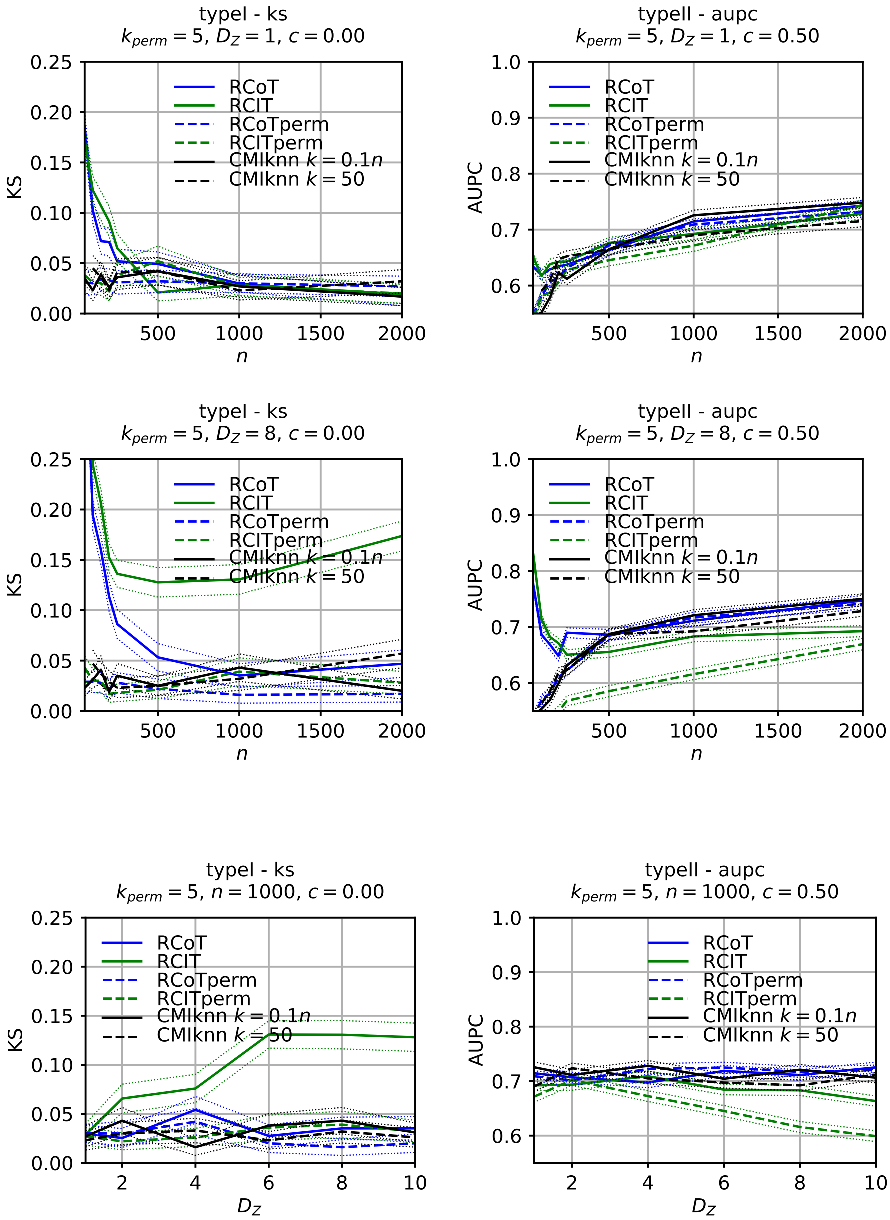

Figure S6: Numerical experiments for a version of model (6) where Z is independent of X and Y when X and Y are dependent under H_1, that is $X = g_X(c\epsilon_b + \epsilon_X)$ and $Y = g_Y(c\epsilon_b + \epsilon_Y)$, which is the setup studied in Zhang et al. (2011); Strobl et al. (2017). Additionally, we show results for the kernel measures combined with the proposed nearest-neighbor permutation test with $k_{\text{perm}} = 5$. The upper two rows depict KS and AUPC for different sample sizes and $D_Z = 1, 8$. The two bottom panels show different dimensions D_Z for $n = 1000$.

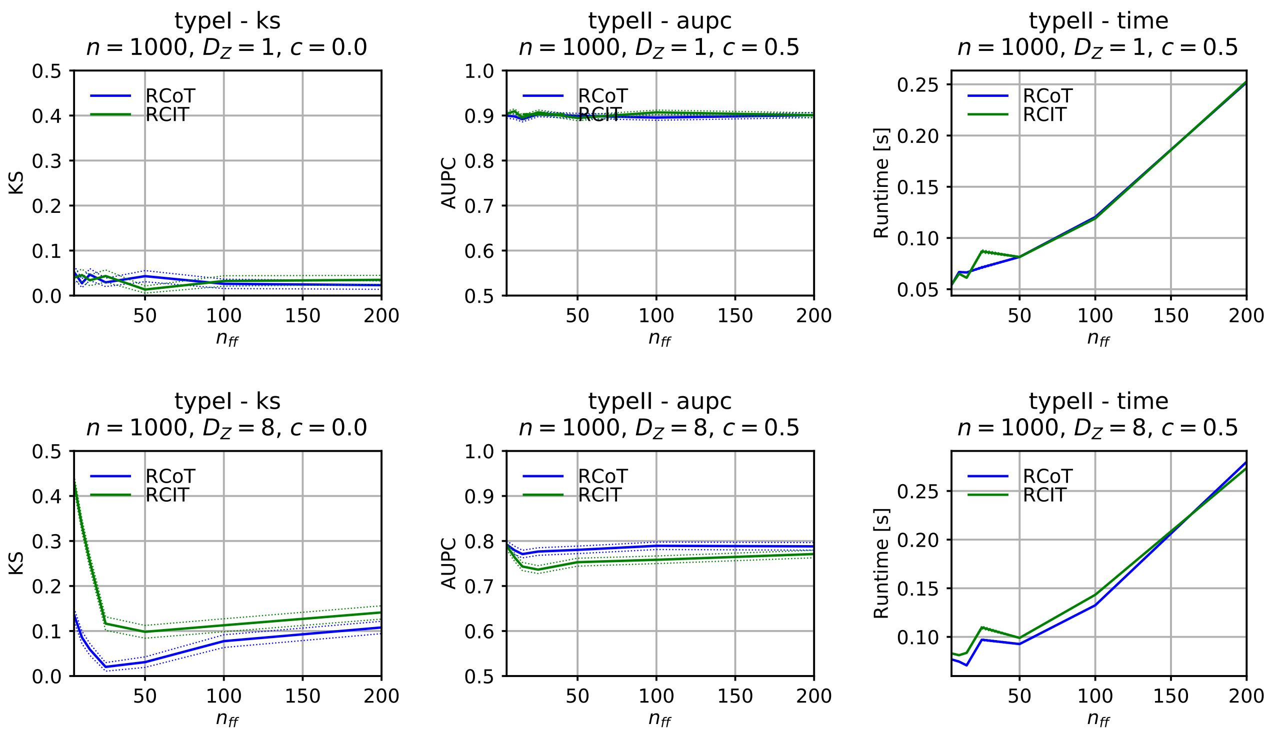

Figure S7: Choice of number of fourier features (n_{ff}) for random fourier-feature based kernel-measures for model (6). Shown are KS (left column), AUPC (center column), and runtime (right column) for a sample size experiment with $D_Z = 1$ (top row) and $D_Z = 8$ (bottom row). n_{ff} corresponds to the number of features in subspace Z, the number of fourier features in subspaces X and Y is fixed to 5 as implemented in `https://github.com/ericstrobl/RCIT`. While for $D_Z = 1$ $n_{\text{ff}} > 10$ yields similar results, for $D_Z = 8$ both the KS and AUPC metrics are more sensitive to the choice of n_{ff}. The default $n_{\text{ff}} = 25$ here gives the most well-calibrated result, but this calibration is quite unstable for smaller or larger n_{ff} values. The runtime of RCIT and RCoT scales roughly quadratically in the number of fourier features.

Table S1: Results from Wang et al. (2015) together with results from RCoT and the CMIknn test. The experiments are described in Wang et al. (2015). Examples 1–4 correspond to conditional independence showing false positives and Examples 5–8 to dependent cases showing true positives at the 5% significance level. CMIknn was run with $k_{\mathrm{CMI}} = 0.2n$ and $k_{\mathrm{perm}} = 5, 10$. The numbers 50..250 denote the sample size.

	Example 1					Example 2				
Test	50	100	150	200	250	50	100	150	200	250
CDIT	0.035	0.034	0.05	0.057	0.048	0.046	0.053	0.055	0.048	0.058
CI.test	0.041	0.051	0.037	0.054	0.041	0.062	0.046	0.044	0.045	0.039
KCI.test	0.039	0.043	0.041	0.04	0.046	0.035	0.004	0.037	0.047	0.05
Rule-of-thumb	0.017	0.027	0.028	0.033	0.033	0.034	0.052	0.044	0.042	0.045
RCoT	0.074	0.059	0.055	0.043	0.050	0.056	0.056	0.069	0.055	0.073
CMIknn ($k_{\mathrm{perm}} = 5$)	0.064	0.055	0.050	0.053	0.045	0.076	0.060	0.074	0.061	0.065
CMIknn ($k_{\mathrm{perm}} = 10$)	0.058	0.061	0.057	0.058	0.046	0.075	0.066	0.053	0.057	0.071
	Example 3					Example 4				
Test	50	100	150	200	250	50	100	150	200	250
CDIT	0.035	0.048	0.055	0.053	0.043	0.049	0.054	0.051	0.058	0.053
CI.test	0.222	0.363	0.482	0.603	0.677	0.043	0.064	0.066	0.05	0.053
KCI.test	0.058	0.047	0.057	0.061	0.054	0.037	0.035	0.058	0.039	0.049
Rule-of-thumb	0.019	0.038	0.032	0.039	0.039	0.037	0.04	0.055	0.059	0.053
RCoT	0.074	0.047	0.046	0.053	0.054	0.115	0.072	0.066	0.061	0.053
CMIknn ($k_{\mathrm{perm}} = 5$)	0.044	0.043	0.046	0.046	0.054	0.084	0.071	0.067	0.079	0.070
CMIknn ($k_{\mathrm{perm}} = 10$)	0.063	0.065	0.061	0.076	0.067	0.101	0.113	0.106	0.098	0.084
	Example 5					Example 6				
Test	50	100	150	200	250	50	100	150	200	250
CDIT	0.898	0.993	1	1	1	0.752	0.995	1	1	1
CI.test	0.978	1	1	1	1	0.468	0.434	0.467	0.476	0.474
KCI.test	0.158	0.481	0.557	0.602	0.742	0.296	0.862	0.995	1	1
Rule-of-thumb	0.368	0.793	0.927	0.983	0.994	1	1	1	1	1
RCoT	0.817	0.986	0.998	1	1	0.301	0.533	0.679	0.807	0.860
CMIknn ($k_{\mathrm{perm}} = 5$)	0.782	0.981	0.998	1	1	0.806	0.997	0.999	1	1
CMIknn ($k_{\mathrm{perm}} = 10$)	0.855	0.995	1	1	1	0.805	0.995	1	1	1
	Example 7					Example 8				
Test	50	100	150	200	250	50	100	150	200	250
CDIT	0.918	0.998	1	1	1	0.361	0.731	0.949	0.977	0.994
CI.test	0.953	0.984	0.983	0.995	0.987	0.456	0.476	0.464	0.461	0.485
KCI.test	0.574	0.947	0.998	1	1	0.089	0.401	0.685	1	1
Rule-of-thumb	0.073	0.302	0.385	0.514	0.515	0.043	0.233	0.551	0.851	0.972
RCoT	0.594	0.880	0.962	0.985	0.991	0.275	0.392	0.470	0.624	0.654
CMIknn ($k_{\mathrm{perm}} = 5$)	0.753	0.963	0.992	0.997	1	0.302	0.644	0.804	0.916	0.958
CMIknn ($k_{\mathrm{perm}} = 10$)	0.798	0.976	0.999	0.999	0.999	0.323	0.680	0.832	0.920	0.971