A ALGORITHMIC COMPARISON

	SDCIT	KCIPT
H_0	$\Omega = \Omega_\pi$	$\{\Omega_i^{(1)} = \Omega_{\pi,i}^{(2)}\}_{i=1}^B$
T	modified MMD	averaged MMDs
null distribution	half-sampling without replacement	aggregated bootstrap null distributions
# of permutations	1 for T $b+1$ for null.	B for T
# of MM(S)Ds	1 for T $b+1$ for null.	B for T Bb for null.

Table 2: Comparison of SDCIT and KCIPT

B TAKING PERMUTATION ERROR INTO ACCOUNT

Our test statistic measures the distance between the original sample (representing P_{xyz}) and a pseudo-null sample (representing $\mathrm{P}_{xz}\mathrm{P}'_{y|z}$), where $\mathrm{P}'_{y|z}$ approximates $\mathrm{P}_{y|z}$. Ideally, the test statistic and its null distribution will be reliably estimated if permutation error is small and, hence, $\mathrm{P}'_{y|z}$ approximates $\mathrm{P}_{y|z}$ well.

We first relate an MMSD estimate and its corresponding permutation error during the estimate, and provide a means to adjust MMSD estimates. Let T be an MMSD estimate given K_{xz}, K_y, and D (see Algorithm 2). Let τ be an MMSD estimate assuming $\mathrm{K}_x = \mathbf{1}_{n\times n}$, that is $\tau = \mathrm{MMSD}(\mathrm{K}_z, \mathrm{K}_y, \mathrm{D})$. In other words, τ is the MMSD estimate between $\mathrm{P}'_{y|z}\mathrm{P}_z$ and $\mathrm{P}_{y|z}\mathrm{P}_z$. While T is, roughly, about the conditional dependence between X and Y given Z, τ measures permutation error, i.e., discrepancy between $(\mathbf{y}, \mathbf{z})$ and $(\pi\mathbf{y}, \mathbf{z})$. We illustrate a null distribution $\{T_i\}_{i=1}^b$ and its associated $\{\tau_i\}_{i=1}^b$ in Figure 7. We can clearly observe that the distribution of τ is centered at 0 but still there are lots of null samples associating non-negligible errors.

We then formulate T (under a permutation error) is the function of unknown T^* (under zero permutation error) and τ. We assume a linear model $T = T^* + \beta\tau + \epsilon$ where ϵ is *assumed* a zero-mean Gaussian noise. Given a null distribution $\{(T_i, \tau_i)\}_{i=1}^b$, we can learn β by fitting a linear model. Then, the null distribution $\{T_i\}_{i=1}^b$ is adjusted to $\{T_i - \beta\tau_i\}_{i=1}^b$ and our test statistic is also adjusted similarly. Such adjustment yields a null distribution with smaller variance as shown in Figure 8. The adjustment slightly improves both power and calibratedness.

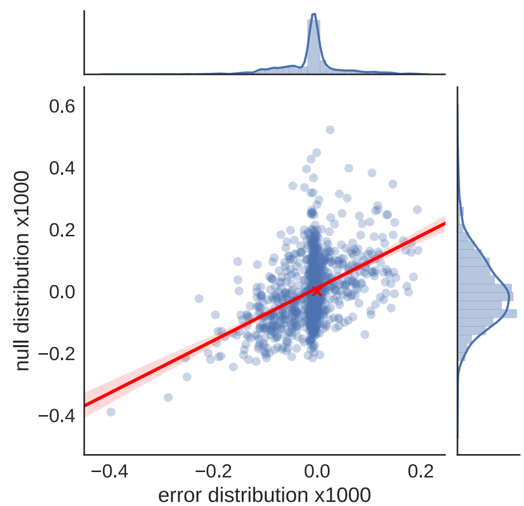

Figure 7: A null distribution and its corresponding errors measured with MMSD. A red cross near origin indicates the test statistic and its corresponding error. A red line indicates a fitted linear model.

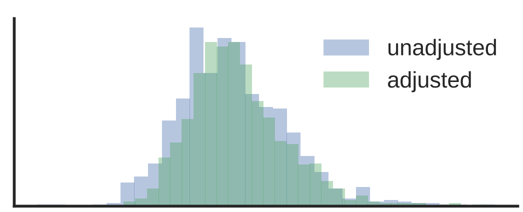

Figure 8: Unadjusted and adjusted null distributions.