1 Synthetic Data Experiments With Random Cov Matrix Per Domain

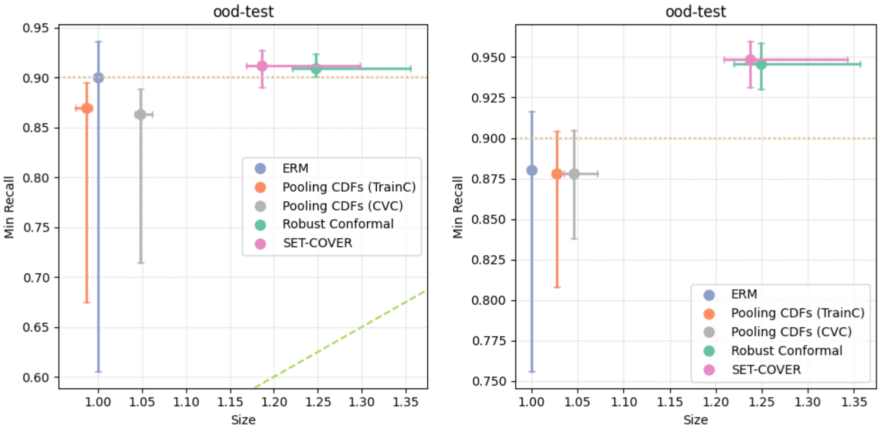
In Theorem 3.4 we have shown a theoretical generalization result, but under the limitation of shared covariance structure across domains (up to a scaling factor). Our results in the synthetic data experiment, presented in Section 5.1 empirically support this result. In this section we want to test whether the generalization to new domains can hold also in DGPs where the covariance between domains does not share exactly same structure. To this end, we recall the DGP presented in Section 5.1:

$$\begin{aligned}
&Z_e \sim U[u_{\text{low}}, u_{\text{high}}] \\
&Y \sim Bernoulli(0.5) \\
&X \sim Y(\mu + Z_e \nu) + N(0, \Sigma)
\end{aligned}$$

In the following experiment we change the covariance matrix to be domain-specific in the following way:

1. We sample for each domain a diagonal matrix, D_e, with diagonal values sampled from a normal distribtuion with $\mu = \sigma$ and $std = 0.05$ (this process generates std values, which are than squared to form the diagonal values of D_e). σ values are the same as set in the original experiment from Section 5.1.
$$D_e = diag([D_{e,1}^2, ..., D_{e,d}^2]$$
$$\forall 1 \leq i \leq d \quad D_{e,i} \sim N(\sigma, 0.05)$$
2. For each domain we sample uniformly a rotation matrix Q_e.
3. For each domain we set the covaraince matrix $\Sigma_e = Q_e^T D_e Q_e$

All other experiments' hyper-parameters are the same as the original experiment from Section 5.1. The results are presented in Figure 1 and Table 1.



(a) 50d

(b) 10d

Figure 1: Each figure represents Min-Recall over Avg Set Size cross. y-axis represents min-recall, and x-axis represents average set size. Each cross shows the median and the 25th and 75th percentiles for both metrics across domain. **Blue** represents ERM predictor, **Orange** represents Pooling CDFs (TrainC), **Grey** represents Pooling CDFs (CVC), **Green** represents robust conformal predictor, and **Pink** represents SET-COVER. The horizontal solid line represents the 90% recall target value, and dashed yellow diagonal line represents performance of a random predictor.

Table 1: OOD Performance on synthetic Datasets

(a) 10d

Model	Median Min Recall ↑	Median Avg Size ↓	Recall ≥ 90% Pctg ↑
ERM	0.88	1.0	0.39
CDF Pooling-(TrainC)	0.87	1.02	0.30
CDF Pooling-(CVC)	0.87	1.04	0.32
Robust-Conformal	0.94	1.24	0.94
SET-COVER	0.94	1.23	0.92

(b) 50d

Model	Median MinRecall ↑	Median Avg Size ↓	Recall ≥ 90% Pctg ↑
ERM	0.90	1.0	0.52
CDF Pooling-(TrainC)	0.86	0.98	0.19
CDF Pooling-(CVC)	0.86	1.04	0.27
Robust-Conformal	0.90	1.24	0.71
SET-COVER	0.91	1.18	0.68