

Figure 1: Synthetic experiments. Mechanisms are randomly instantiated. Task is to classify target  $> 0$ .

Table 1: Summary of tasks and their associated distribution shifts.

Task	Covariate shift (OTDD)	Concept shift (FDD)	Label shift (L2 distance)
Food Stamps	14.20	640.82	0.0008
Income	30.60	1.40	0.0060
Public Coverage	5.79	4.06	0.1701
Unemployment	75.47	13,389,512.51	0.0003
ANES	13.60	2.23	0.0025
Diabetes	12.28	0.10	0.0332
Hypertension	4.69	0.04	0.0022
Hospital Readmission	42.37	1.30	0.0060
Childhood Lead	1.30	0.01	0.0026
Sepsis	6609.73	8.44	0.0040
ICU Length of Stay	56,439,324,672.00	47,042,729,585.25	0.0033
ICU Hospital Mortality	64,479,092,736.00	42,639,188,407.47	0.0015
ASSISTments	24,054.59	1137.42	0.0670
College Scorecard	43,566.39	2116.63	0.0337
SIPP	6,344,306.0	5,752,406.89	0.0751
MEPS	66.28	4.01	0.0013

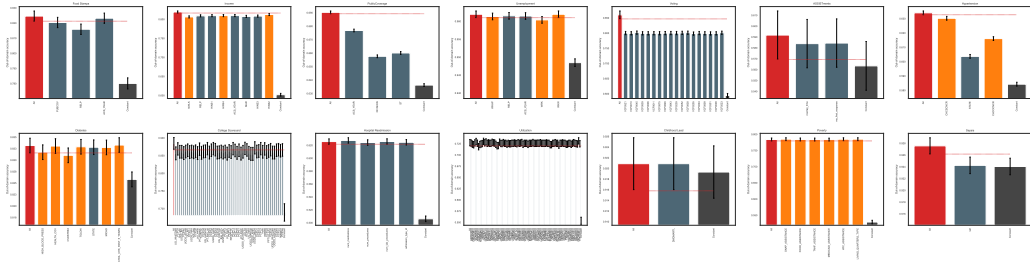


Figure 2: Removing one feature at a time. Anti-causal features are colored in orange, non-causal in grey.