Table 1: Computations on the $\mathbb{I}$ dataset: p-values for FILO2^x vs FILO2 on the left and FILO2^x (long) vs FILO2 (long) on the right.

p-values in bold are associated with rejected hypothesis when $\bar{\alpha} = 0.002778$.

The last column of each group contains a p-value interpretation. In particular, FILO2^x is not statistically different from FILO2 when H_0 cannot be rejected (Similar), FILO2^x is statistically better when both H_0 and H_1 are rejected (Better), and, finally, FILO2^x is statistically worse when H_0 is rejected and H_1 is not rejected (Worse).

| Solvers | FILO2^x vs FILO2 | | FILO2^x (long) vs FILO2 (long) | |
	$H_0 \,/\, H_1$	Outcome	$H_0 \,/\, H_1$	Outcome
2	0.674223 / 0.337111	Similar	0.277355 / 0.869451	Similar
3	0.985435 / 0.492718	Similar	0.082550 / 0.962074	Similar
4	0.570597 / 0.727062	Similar	0.164957 / 0.923177	Similar
5	0.956329 / 0.478165	Similar	0.701181 / 0.662889	Similar
6	0.898317 / 0.565256	Similar	0.142906 / 0.933637	Similar
7	0.985435 / 0.492718	Similar	0.311794 / 0.852874	Similar
8	0.927279 / 0.463639	Similar	0.348810 / 0.835009	Similar
9	0.570597 / 0.727062	Similar	0.294252 / 0.861322	Similar
10	0.498009 / 0.762547	Similar	0.521673 / 0.750996	Similar