Brier score model by model benchmark

Legend: Isotonic regression, TS (Guo et al.), TS (probmetrics), TS (ours), Affine scaling, Quadratic scaling