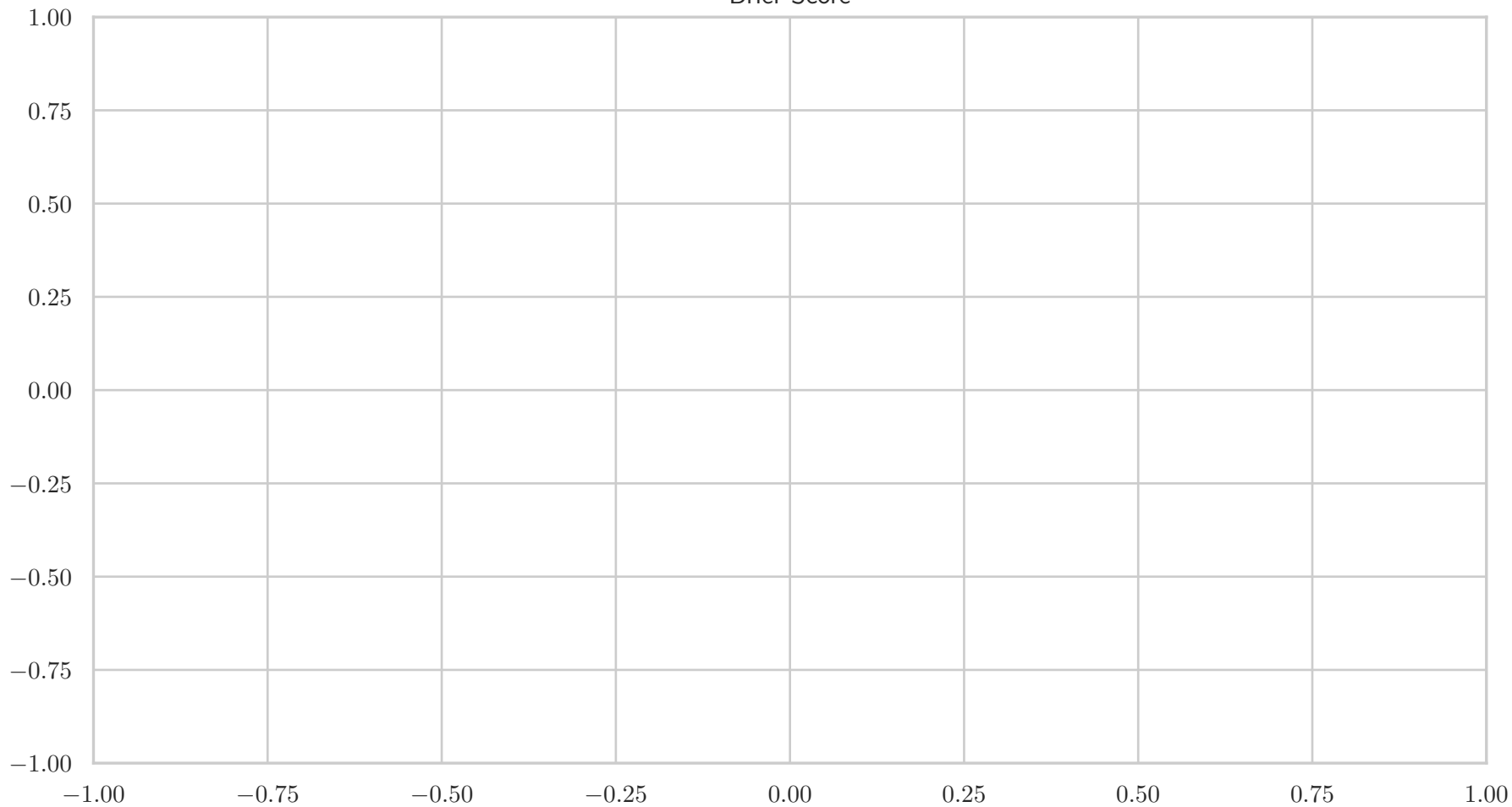


PubMedQA
Brier Score

Per-Sample Correlation(Compute, Score) for
Pythia 12B Param. (Data Scaling)



Per-Sample Correlation(Compute, Score) for
LLM360 Amber 7B Tokens (Param Scaling)