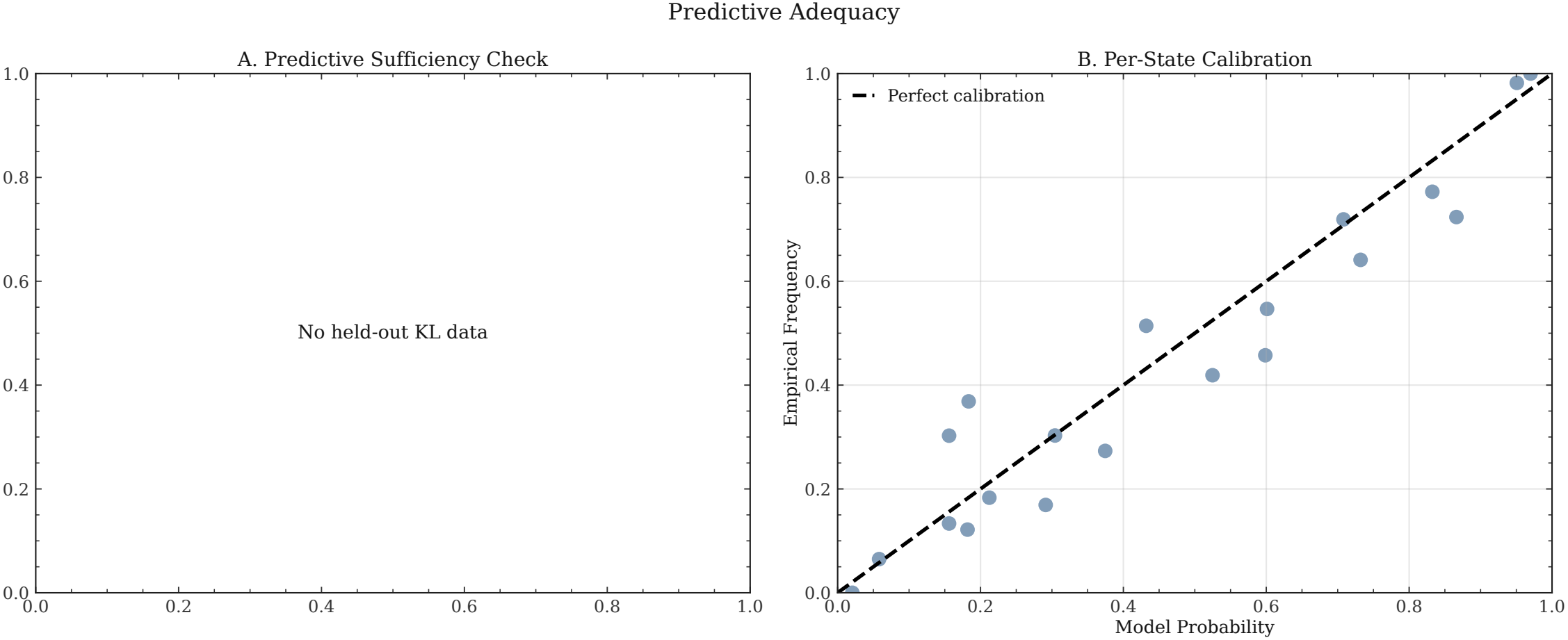
Predictive Adequacy
A. Predictive Sufficiency Check
No held-out KL data
B. Per-State Calibration
Perfect calibration
Empirical Frequency
Model Probability
0.0
0.2
0.4
0.6
0.8
1.0