Figure 4: Calibration Analysis — Expected Calibration Error (N=120, K=8)

(a) Reliability Diagram (N=120)

ECE = 0.200

(b) Calibration Gap per Bin

Overconfident
Underconfident
ECE = 0.207