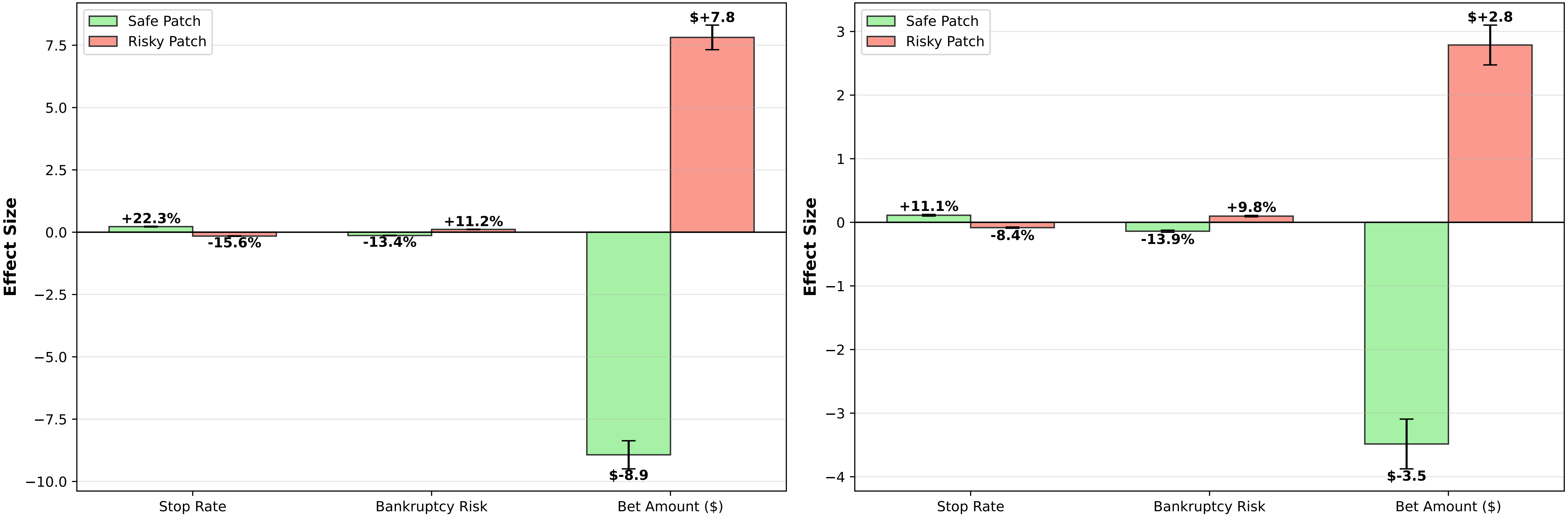
Comprehensive Patching Effects: All Metrics (Real Data)
149 Causal Features from Activation Patching Experiment

Safe Context ($140 balance)
Patching Effects on All Metrics

Risky Context ($20 balance)
Patching Effects on All Metrics



Real experimental data: 149 causal features from GPU 4 & 5
Error bars: Standard Error of Mean (SEM)
Stop Rate: bet = $0, Bankruptcy Risk: bet > balance×50%, Bet Amount: average bet