ASD vs ASE for Consultancy [consultant starts, 2 turns] (brier)

r = 0.456 (p = 0.440)

ASD vs ASE for Debate [simultaneous, 4 turns] (brier)

r = 0.819 (p = 0.181)

ASD vs ASE for Blind (brier)
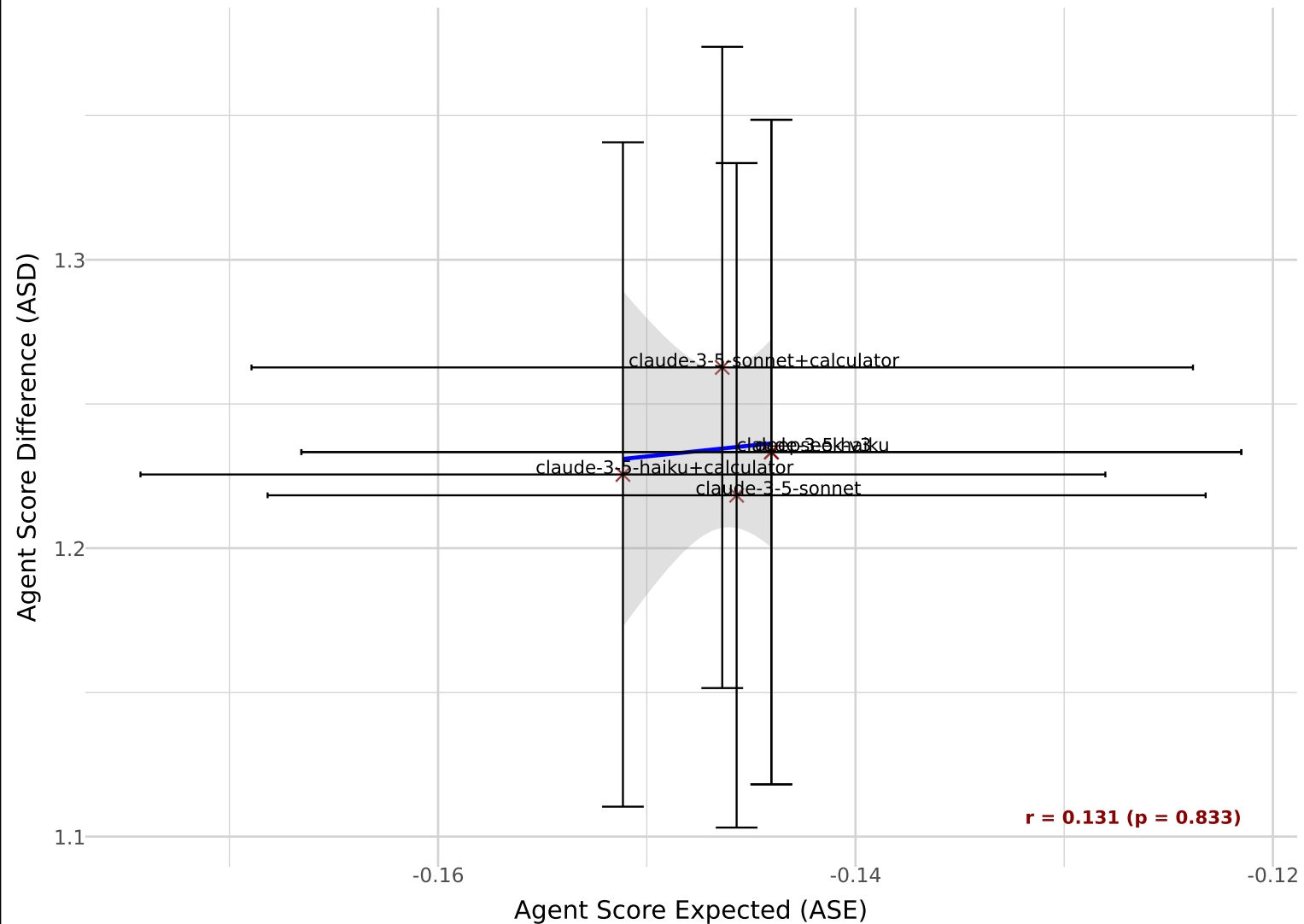
ASD vs ASE for Debate [sequential, 4 turns] (brier)

ASD vs ASE for Consultancy [client starts, turns] (brier)

r = -0.660 (p = 0.226)
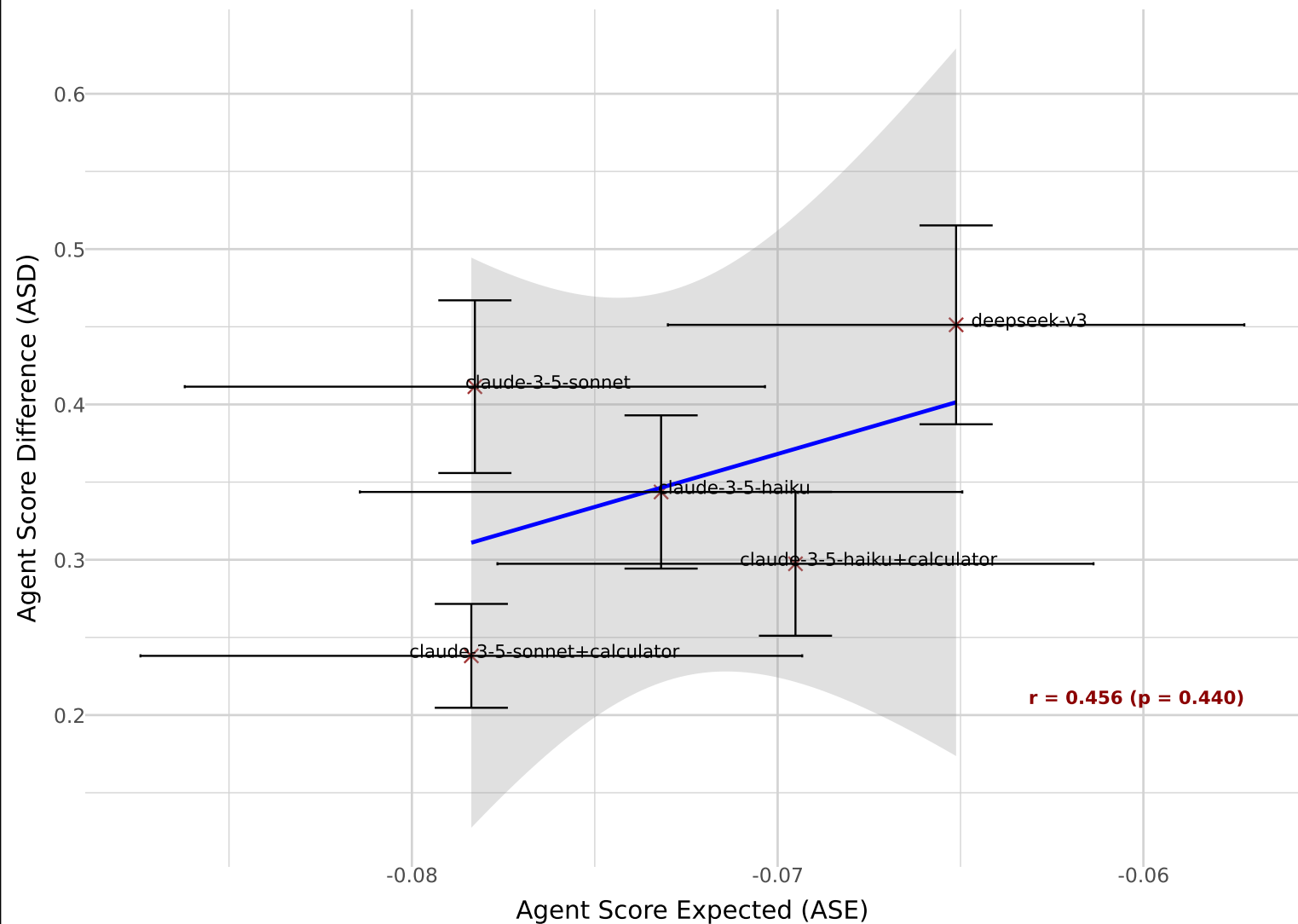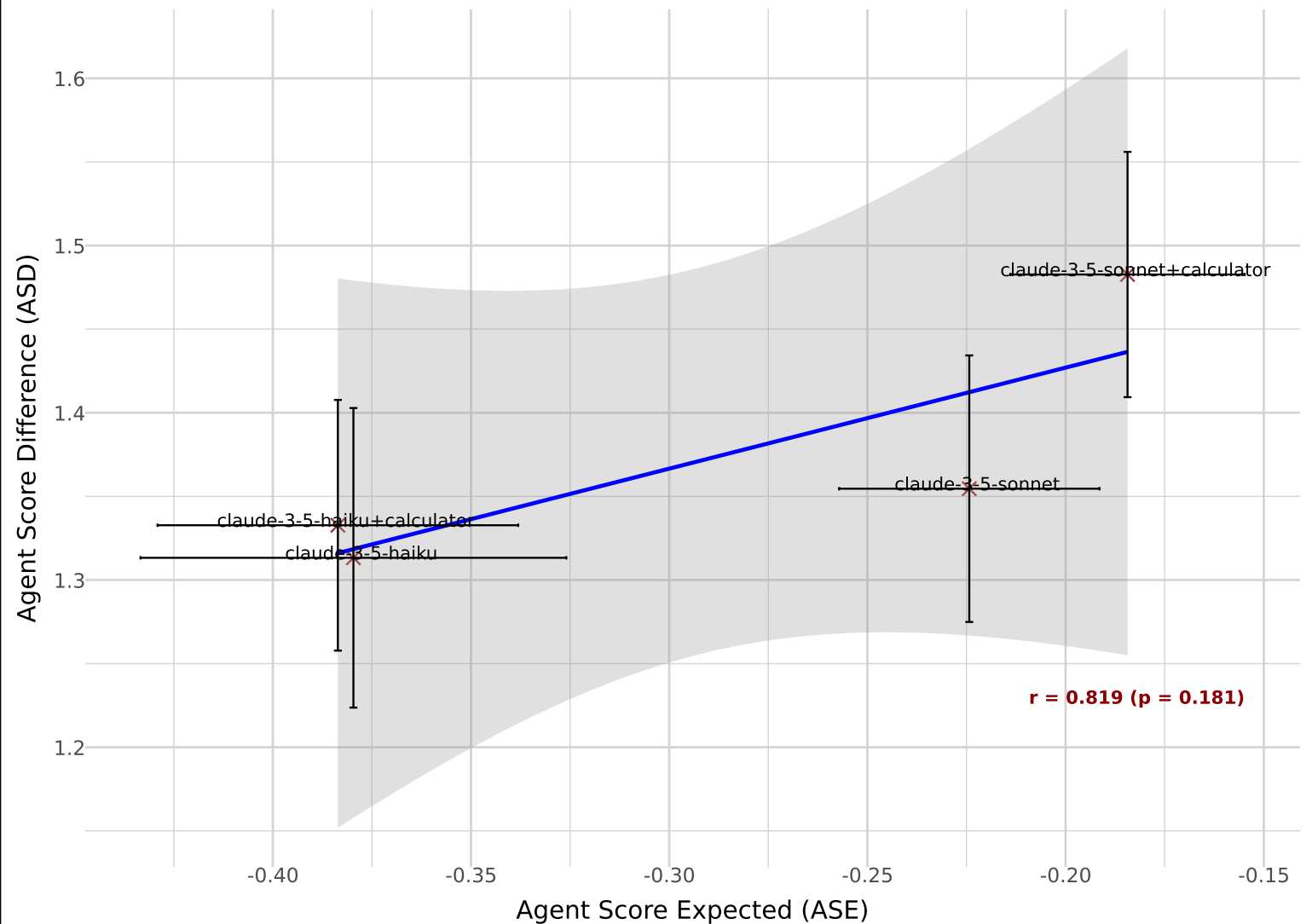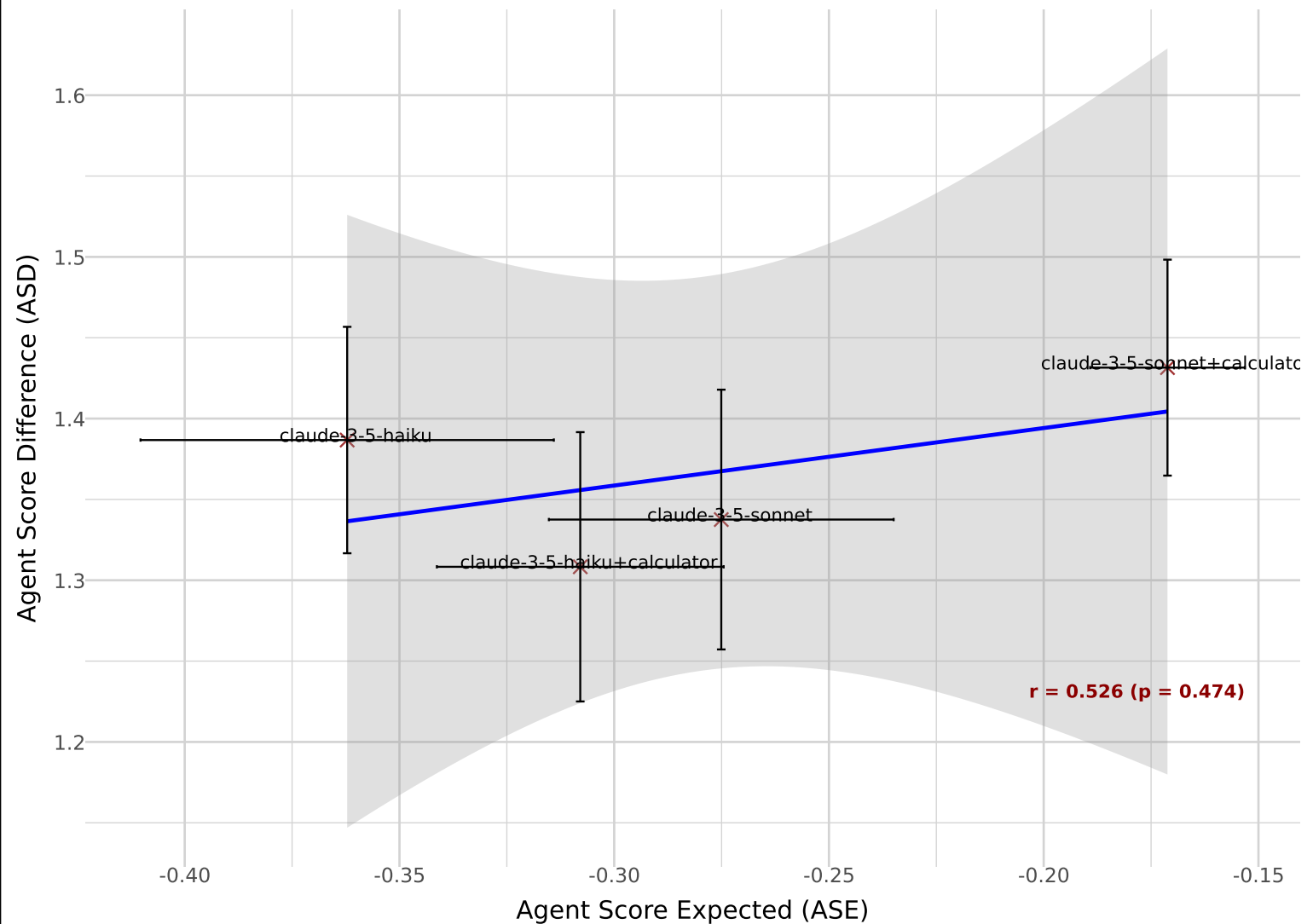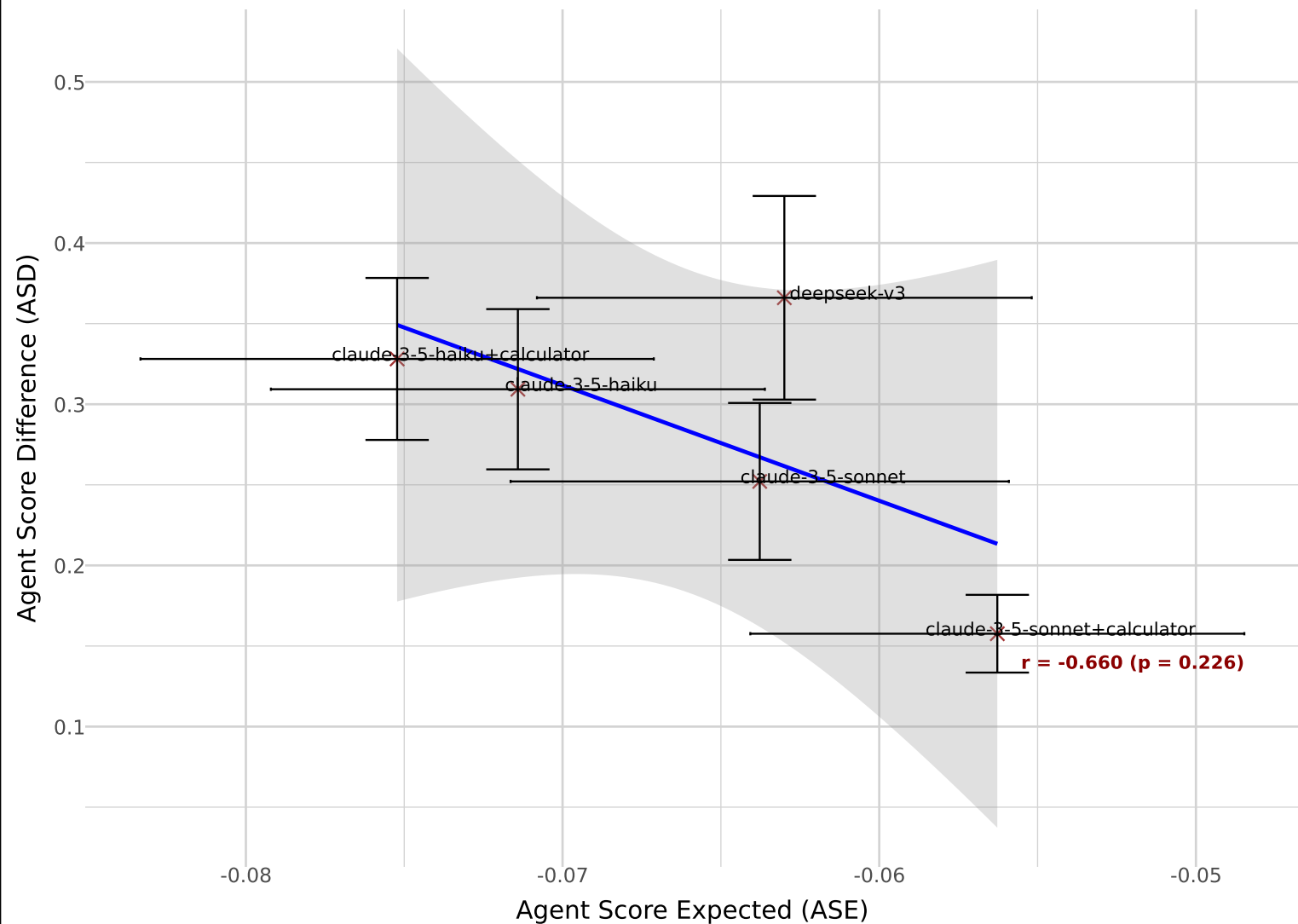
ASD vs ASE for Consultancy [client starts, 2 turns] (brier)

ASD vs ASE for Debate [simultaneous, 2 turns] (brier)

r = -0.279 (p = 0.721)

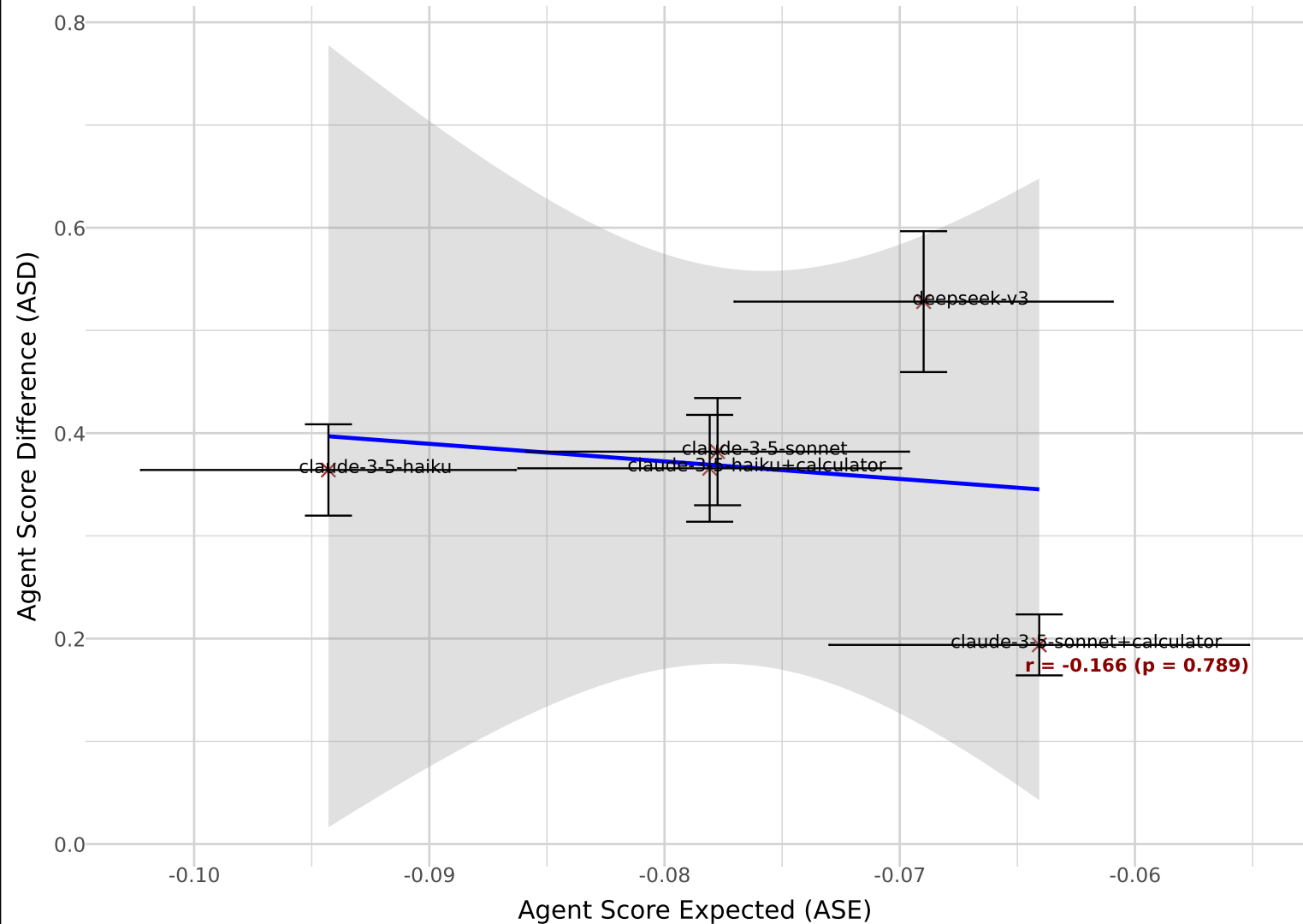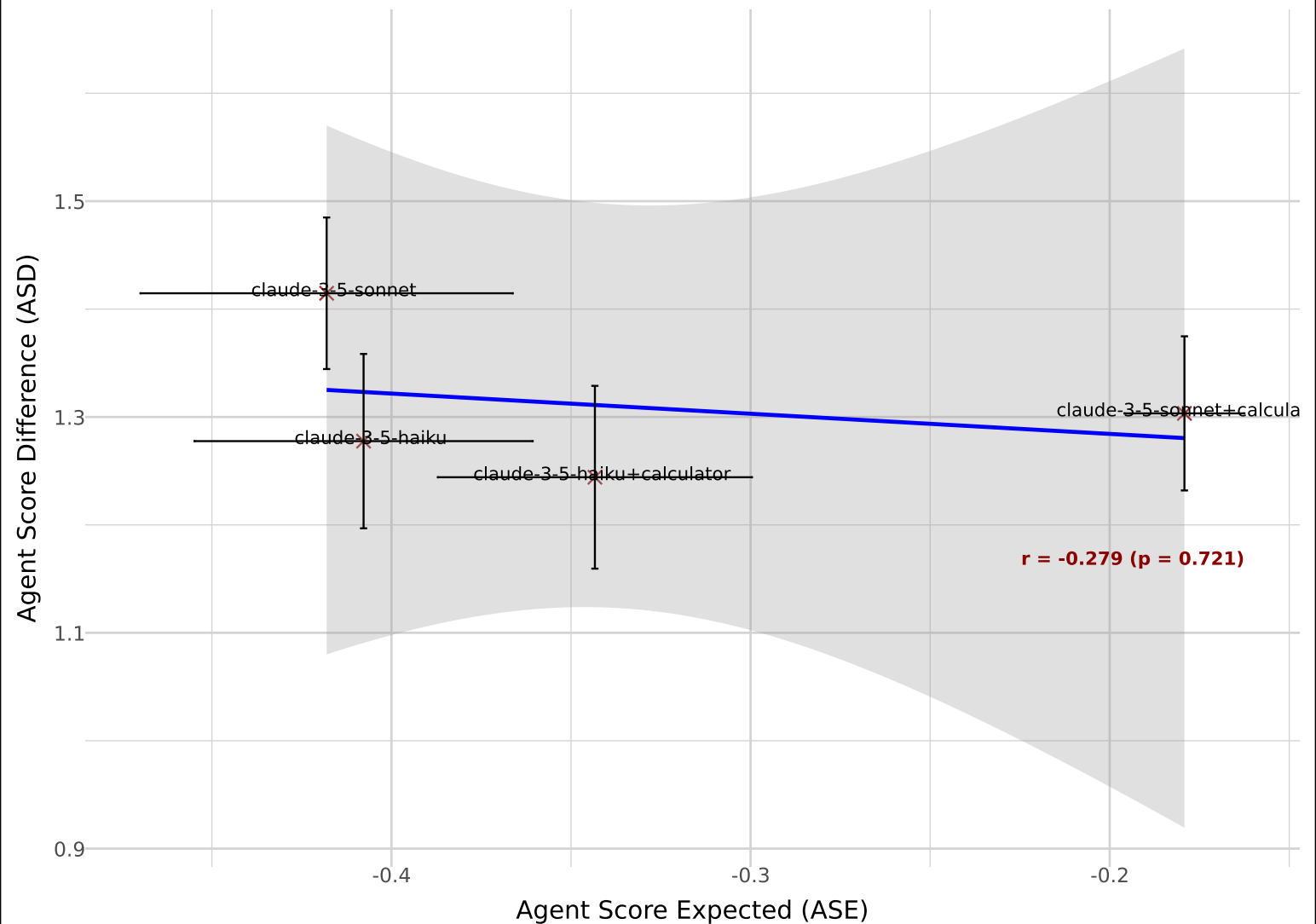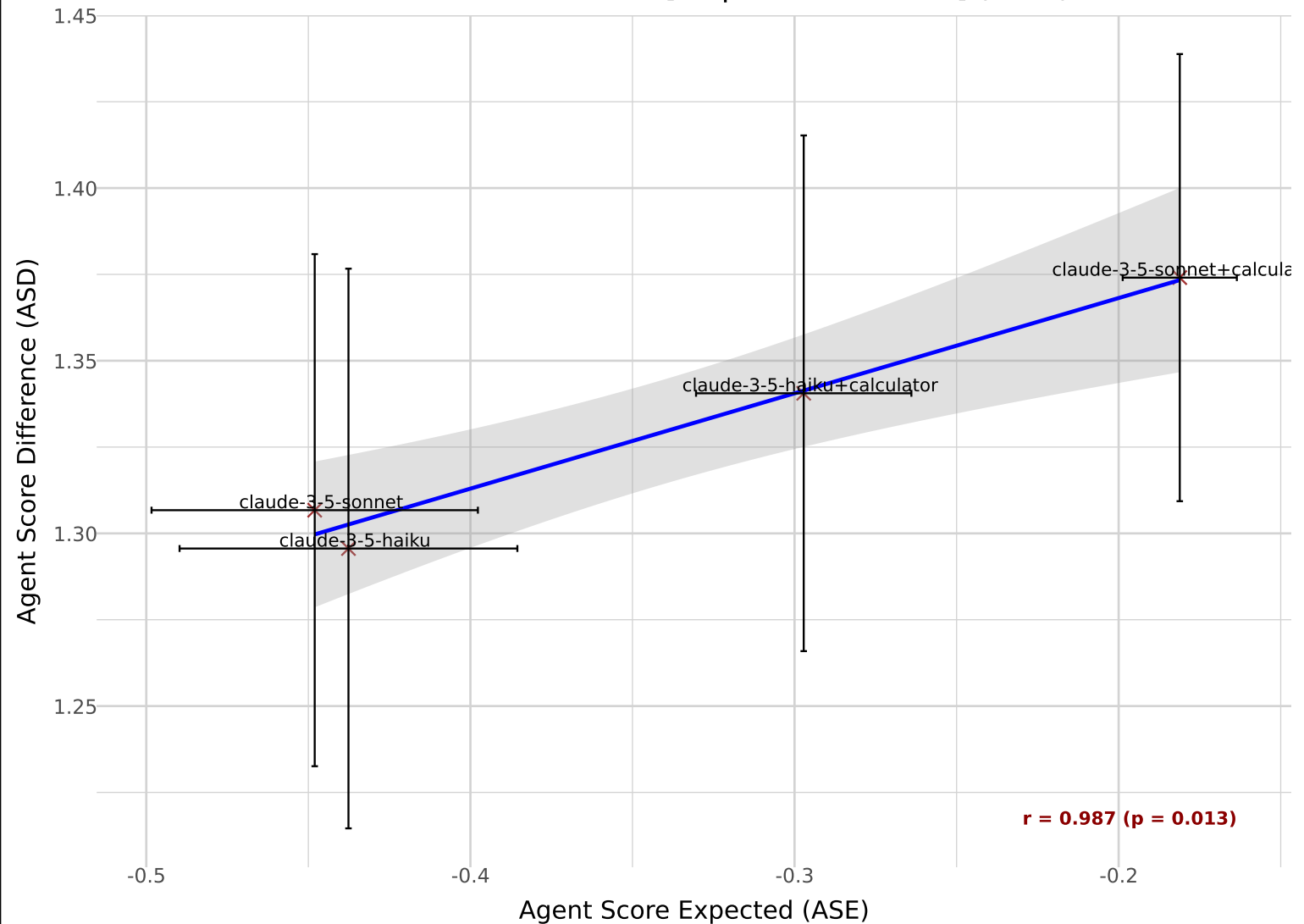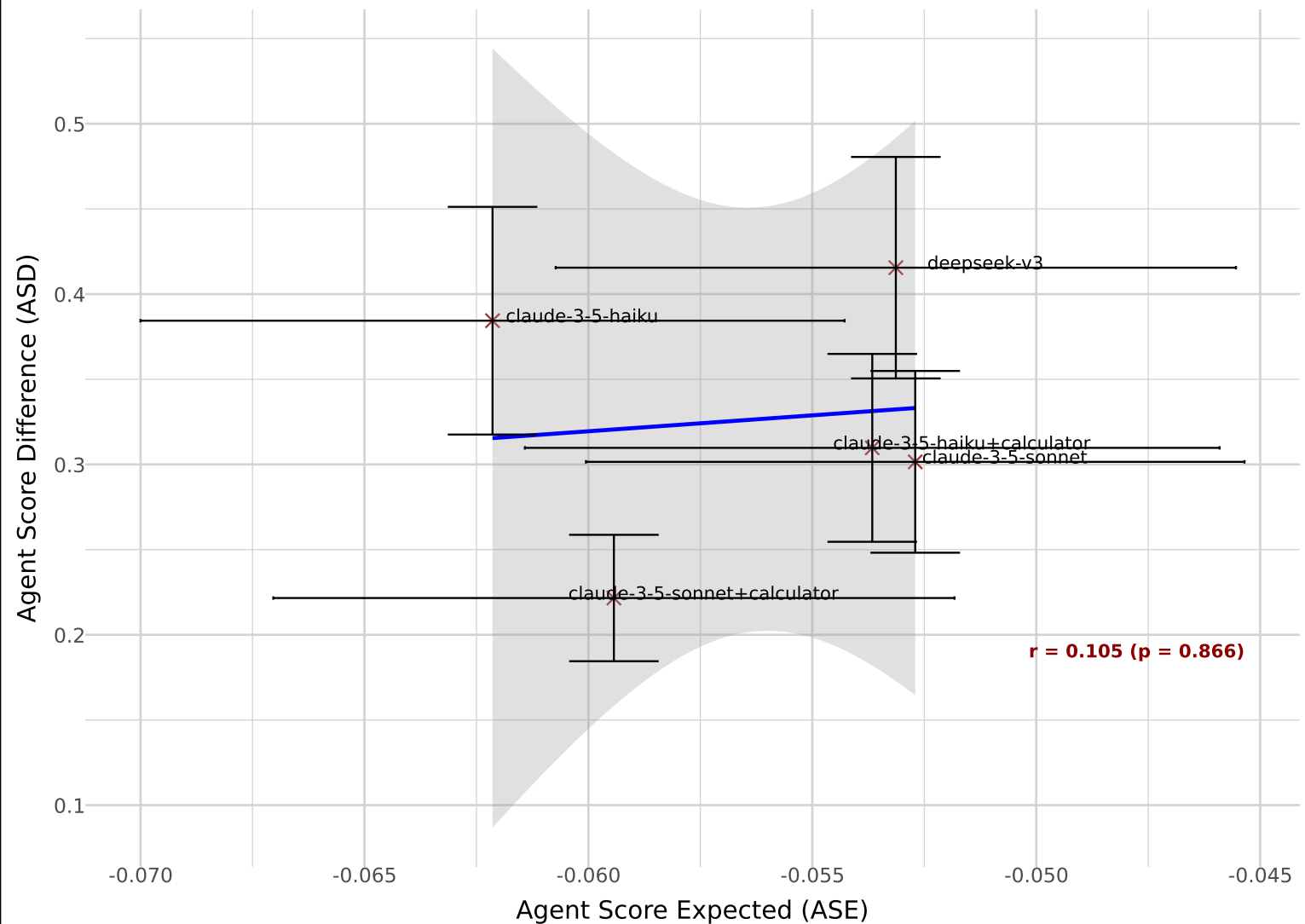ASD vs ASE for Debate [sequential, 2 turns] (brier)

r = 0.987 (p = 0.013)

ASD vs ASE for Consultancy [consultant starts, 4 turns] (brier)

r = 0.105 (p = 0.866)

deepseek-v3

claude-3-5-haiku

claude-3-5-haiku+calculator

claude-3-5-sonnet

claude-3-5-sonnet+calculator

Agent Score Expected (ASE)

Agent Score Difference (ASD)

ASD vs ASE for Propaganda (brier)