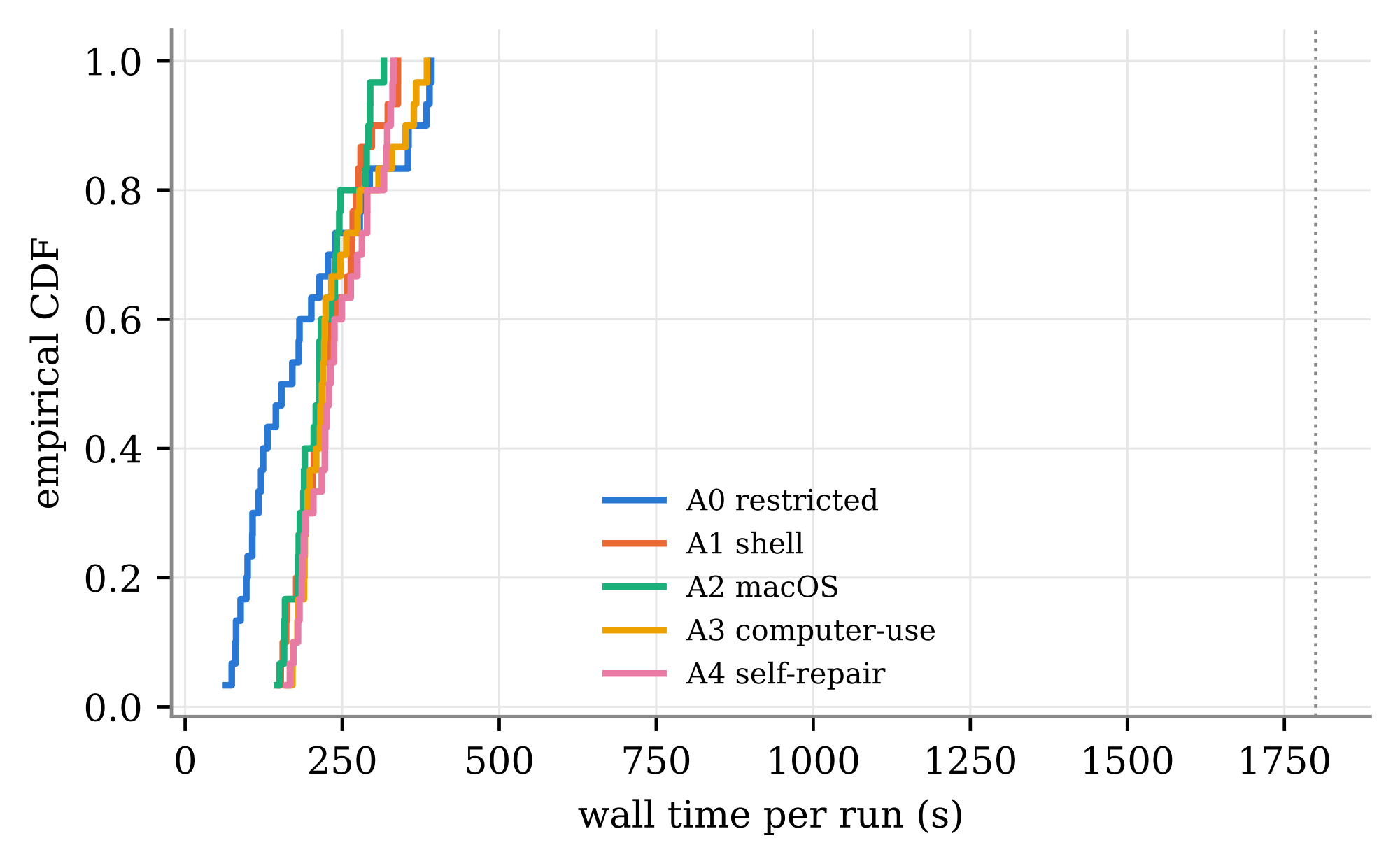

empirical CDF
1.0
0.8
0.6
0.4
0.2
0.0
0
250
500
750
1000
1250
1500
1750
wall time per run (s)
A0 restricted
A1 shell
A2 macOS
A3 computer-use
A4 self-repair