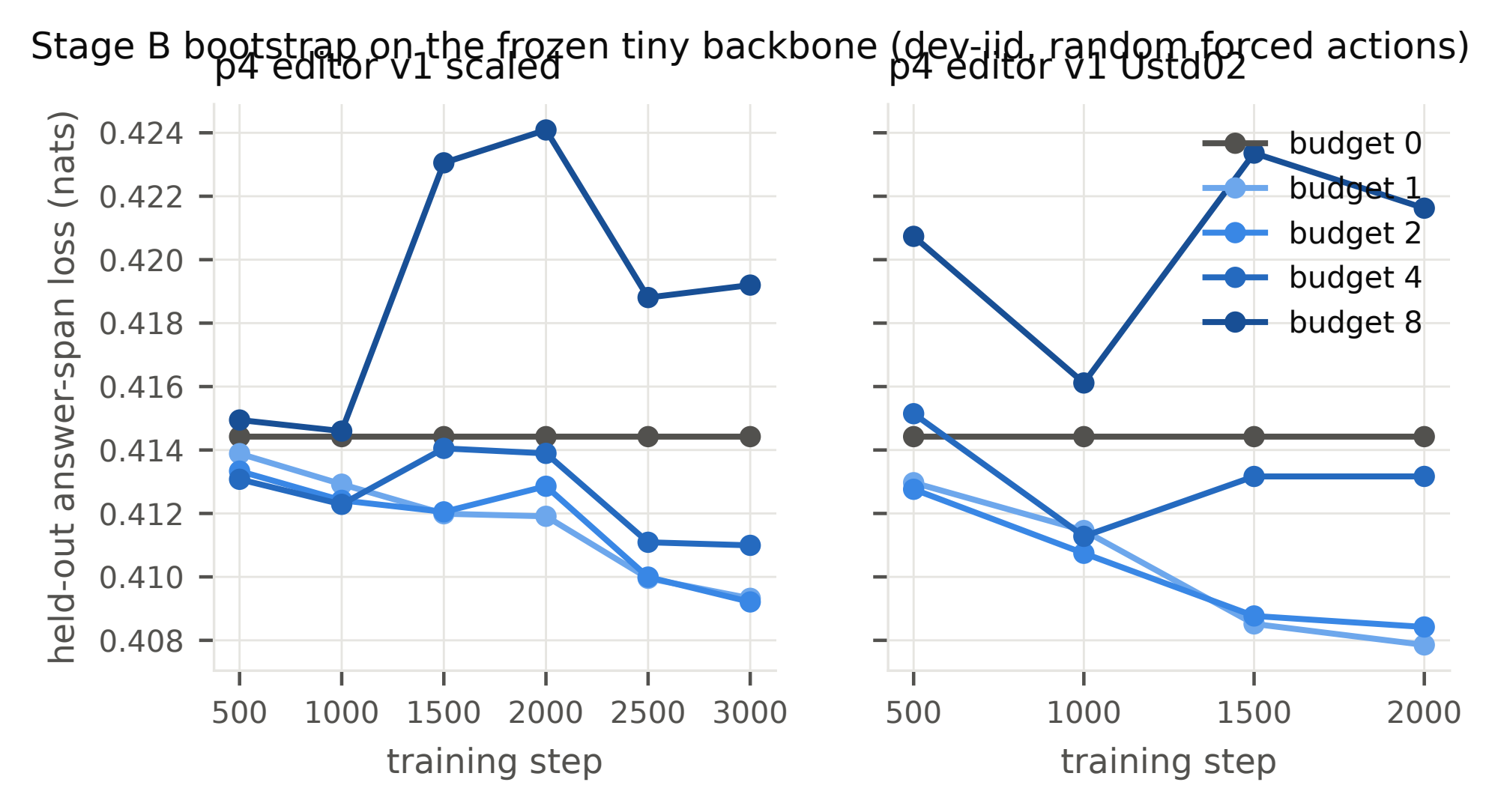

Stage B bootstrap on the frozen tiny backbone (dev-iid, random forced actions)
p4 editor v1 scaled
p4 editor v1 Ustd02
held-out answer-span loss (nats)
0.424
0.422
0.420
0.418
0.416
0.414
0.412
0.410
0.408
500
1000
1500
2000
2500
3000
training step
budget 0
budget 1
budget 2
budget 4
budget 8