In Monte Carlo Exploring Starts Algorithm, It initialized a policy first, and a Q depends on each states and actions, and run the loop forever for each episode to estimating pi approximate equal to pi star. The subloop in this process, Loop for each step of episode, t = T −1, T −2, . . . , 0, what if we give this loop a limit? We do not let it loop until termination, instead, what if we make it limit steps? Like t = T −1, T −2, . . . , a small constant, or t = T −1, T −2, . . .T - 10? Will the pair St,At appears in S0,A0, S1,A1 . . . , St−1,At−1 like old ways? What’s going to happen to S0,A0...until Sc,Ac? And, after run out this algorithm, can we have the same result which is pi approximate equal to pi star? And what does the graph should like?

In each episode, I do not think the estimate value can still approximate to 1.0; Along with the loop running, the fluctuate will be as similar as origin figure. However, the termination will be early so that it will never reach 1.0 and we can not have pi approximate equal pi star.