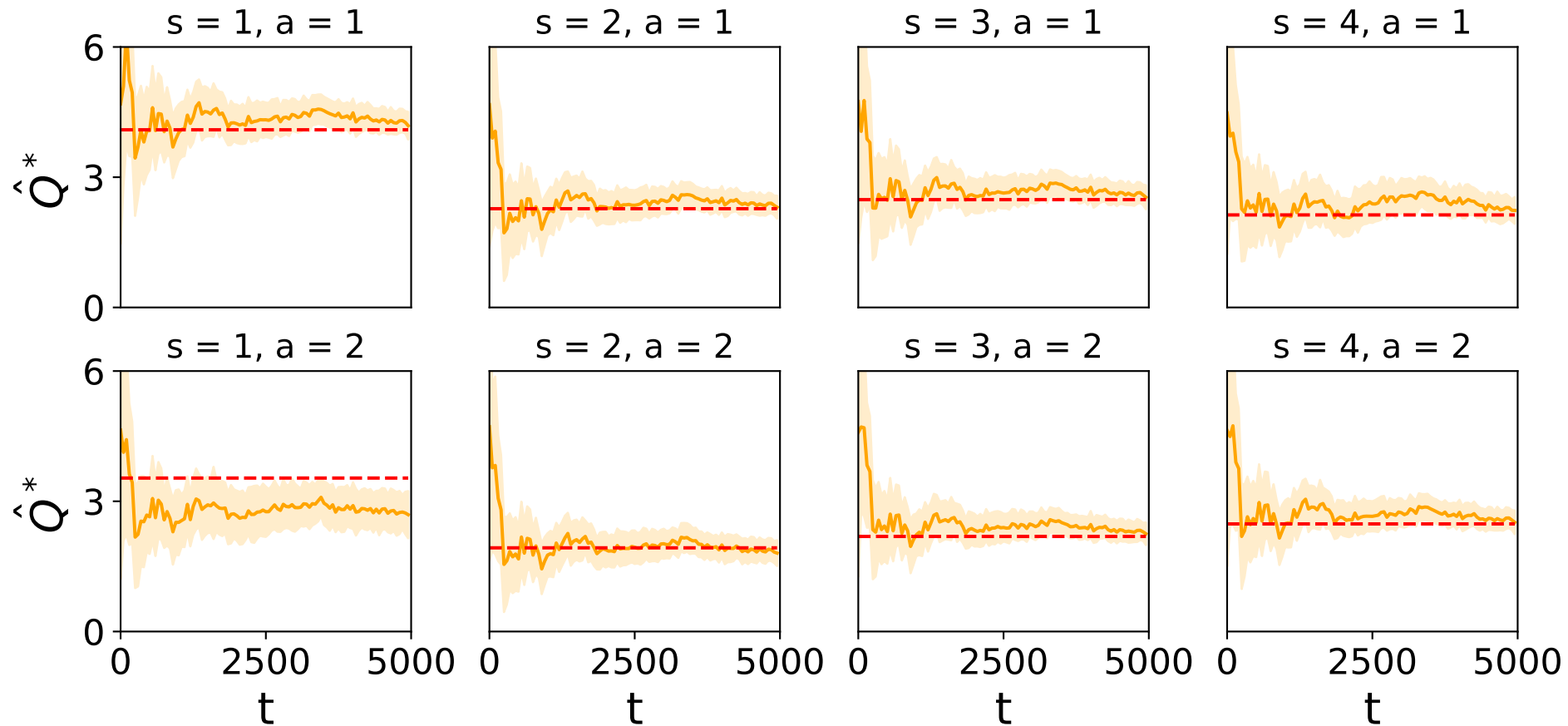


PSRL posterior on PriorMDP ($N_s = 4, N_a = 2$)



— $\mathbb{E}_{\theta_{\mathcal{T}}, \theta_{\mathcal{R}}} [\hat{Q}_{s,a}^*]$

■ $\text{Var}_{\theta_{\mathcal{T}}, \theta_{\mathcal{R}}} [\hat{Q}_{s,a}^*]^{1/2}$

- - - $Q_{s,a}^*$