ζ-R-max optimal policy - last 500 steps