

为何结果不好？为何这么快？ CV？

```
00:17, 04/08; num of cores:16

final_sd_R_trend for[0, 5, 10, 15, 20] the same

Basic setting:[T, rep_times, sd_0, sd_D, sd_u_0, w_0, w_A, [M_in_R, mean_reversion, pois0, u_0_u_D], sd_R_range, t_func] = [None, 48, None, None, 30, 0.5, 1, [True, False, True, 10], [0, 5, 10, 15, 20], None]

=====
[pattern_seed, day, sd_R] = [2, 7, 0]

max(u_0) = 168.8
0_threshold = 80
number of reward locations: 15
0_threshold = 90
number of reward locations: 12
0_threshold = 100
number of reward locations: 9
0_threshold = 110
number of reward locations: 6
target 1 in 4 DONE!
target 2 in 4 DONE!
target 3 in 4 DONE!
target 4 in 4 DONE!
~[["A"

-----
Value of Behaviour policy:57.729
0_threshold = 80
MC for this TARGET:[68.355, 0.062]
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.94, -1.43, -1.07]][[-68.36, -68.36, -10.63]]
std:[[0.44, 0.48, 0.24]][[0.0, 0.0, 0.19]]
MSE:[1.04, 1.51, 1.1][[68.36, 68.36, 10.63]]
MSE(-DR):[[0.0, 0.47, 0.06]][[67.32, 67.32, 9.59]]
=====
0_threshold = 90
MC for this TARGET:[66.714, 0.067]
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.56, -1.02, -0.77]][[-66.71, -66.71, -8.99]]
std:[[0.31, 0.35, 0.23]][[0.0, 0.0, 0.19]]
MSE:[0.64, 1.08, 0.8][[66.71, 66.71, 8.99]]
MSE(-DR):[[0.0, 0.44, 0.16]][[66.07, 66.07, 8.35]]
=====
0_threshold = 100
MC for this TARGET:[66.948, 0.071]
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.27, -3.66, -3.16]][[-66.95, -66.95, -9.22]]
std:[[0.34, 0.37, 0.24]][[0.0, 0.0, 0.19]]
MSE:[3.29, 3.68, 3.17][[66.95, 66.95, 9.22]]
MSE(-DR):[[0.0, 0.39, -0.12]][[63.66, 63.66, 5.93]]
=====
0_threshold = 110
MC for this TARGET:[65.968, 0.07]
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.82, -6.16, -5.36]][[-65.97, -65.97, -8.24]]
std:[[0.4, 0.43, 0.34]][[0.0, 0.0, 0.19]]
MSE:[5.83, 6.17, 5.37][[65.97, 65.97, 8.24]]
MSE(-DR):[[0.0, 0.34, -0.46]][[60.14, 60.14, 2.41]]
=====
[[ 1.04  1.51  1.1  68.36 68.36 10.63]
 [ 0.64  1.08  0.8  66.71 66.71  8.99]
 [ 3.29  3.68  3.17 66.95 66.95  9.22]
 [ 5.83  6.17  5.37 65.97 65.97  8.24]]

time spent until now: 93.9 mins
```

```
-----  
[pattern_seed, day, sd_R] = [2, 7, 5]
```

```
max(u_0) = 168.8  
0_threshold = 80  
number of reward locations: 15  
0_threshold = 90  
number of reward locations: 12  
0_threshold = 100  
number of reward locations: 9  
0_threshold = 110  
number of reward locations: 6  
target 1 in 4 DONE!  
target 2 in 4 DONE!  
target 3 in 4 DONE!  
target 4 in 4 DONE!
```

```
-----  
Value of Behaviour policy:57.738  
0_threshold = 80  
MC for this TARGET:[68.363, 0.083]  
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]  
bias:[[-0.92, -1.43, -1.04]][[-68.36, -68.36, -10.62]]  
std:[[0.55, 0.58, 0.3]][[0.0, 0.0, 0.19]]  
MSE:[1.07, 1.54, 1.08]][[68.36, 68.36, 10.62]]  
MSE(-DR):[[0.0, 0.47, 0.01]][[67.29, 67.29, 9.55]]
```

```
***
```

```
=====
```

```
0_threshold = 90  
MC for this TARGET:[66.722, 0.092]  
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]  
bias:[[-0.56, -1.02, -0.75]][[-66.72, -66.72, -8.98]]  
std:[[0.36, 0.38, 0.27]][[0.0, 0.0, 0.19]]  
MSE:[0.67, 1.09, 0.8]][[66.72, 66.72, 8.98]]  
MSE(-DR):[[0.0, 0.42, 0.13]][[66.05, 66.05, 8.31]]
```

```
***
```

```
=====
```

```
0_threshold = 100  
MC for this TARGET:[66.956, 0.097]  
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]  
bias:[[-3.27, -3.65, -3.14]][[-66.96, -66.96, -9.22]]  
std:[[0.34, 0.37, 0.25]][[0.0, 0.0, 0.19]]  
MSE:[3.29, 3.67, 3.15]][[66.96, 66.96, 9.22]]  
MSE(-DR):[[0.0, 0.38, -0.14]][[63.67, 63.67, 5.93]]
```

```
**
```

```
=====
```

```
0_threshold = 110  
MC for this TARGET:[65.976, 0.097]  
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]  
bias:[[-5.81, -6.14, -5.37]][[-65.98, -65.98, -8.24]]  
std:[[0.43, 0.45, 0.36]][[0.0, 0.0, 0.19]]  
MSE:[5.83, 6.16, 5.38]][[65.98, 65.98, 8.24]]  
MSE(-DR):[[0.0, 0.33, -0.45]][[60.15, 60.15, 2.41]]
```

```
**
```

```
=====
```

```
[[ 1.04  1.51  1.1  68.36 68.36 10.63]  
[ 0.64  1.08  0.8  66.71 66.71  8.99]  
[ 3.29  3.68  3.17 66.95 66.95  9.22]  
[ 5.83  6.17  5.37 65.97 65.97  8.24]]
```

```
[[ 1.07  1.54  1.08 68.36 68.36 10.62]  
[ 0.67  1.09  0.8  66.72 66.72  8.98]  
[ 3.29  3.67  3.15 66.96 66.96  9.22]  
[ 5.83  6.16  5.38 65.98 65.98  8.24]]
```

```
time spent until now: 188.1 mins
```

```
[pattern_seed, day, sd_R] = [2, 7, 10]
```

```
max(u_0) = 168.8
0_threshold = 80
number of reward locations: 15
0_threshold = 90
number of reward locations: 12
0_threshold = 100
number of reward locations: 9
0_threshold = 110
number of reward locations: 6
target 1 in 4 DONE!
target 2 in 4 DONE!
target 3 in 4 DONE!
target 4 in 4 DONE!
```

```
-----
Value of Behaviour policy:57.748
```

```
0_threshold = 80
```

```
MC for this TARGET:[68.371, 0.129]
```

```
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.89, -1.37, -1.0]] [[-68.37, -68.37, -10.62]]
std:[[0.71, 0.74, 0.41]] [[0.0, 0.0, 0.21]]
MSE:[1.14, 1.56, 1.08] [[68.37, 68.37, 10.62]]
MSE(-DR):[[0.0, 0.42, -0.06]] [[67.23, 67.23, 9.48]]
**
```

```
=====
0_threshold = 90
```

```
MC for this TARGET:[66.73, 0.139]
```

```
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.52, -0.97, -0.73]] [[-66.73, -66.73, -8.98]]
std:[[0.53, 0.53, 0.41]] [[0.0, 0.0, 0.21]]
MSE:[0.74, 1.11, 0.84] [[66.73, 66.73, 8.98]]
MSE(-DR):[[0.0, 0.37, 0.1]] [[65.99, 65.99, 8.24]]
***
```

```
=====
0_threshold = 100
```

```
MC for this TARGET:[66.964, 0.143]
```

```
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.23, -3.58, -3.14]] [[-66.96, -66.96, -9.22]]
std:[[0.44, 0.47, 0.32]] [[0.0, 0.0, 0.21]]
MSE:[3.26, 3.61, 3.16] [[66.96, 66.96, 9.22]]
MSE(-DR):[[0.0, 0.35, -0.1]] [[63.7, 63.7, 5.96]]
**
```

```
=====
0_threshold = 110
```

```
MC for this TARGET:[65.984, 0.143]
```

```
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.84, -6.14, -5.4]] [[-65.98, -65.98, -8.24]]
std:[[0.61, 0.64, 0.44]] [[0.0, 0.0, 0.21]]
MSE:[5.87, 6.17, 5.42] [[65.98, 65.98, 8.24]]
MSE(-DR):[[0.0, 0.3, -0.45]] [[60.11, 60.11, 2.37]]
**
```

```
=====
```

```
[ [ 1.04  1.51  1.1  68.36 68.36 10.63]
  [ 0.64  1.08  0.8  66.71 66.71  8.99]
  [ 3.29  3.68  3.17 66.95 66.95  9.22]
  [ 5.83  6.17  5.37 65.97 65.97  8.24]]
```

```
[ [ 1.07  1.54  1.08 68.36 68.36 10.62]
  [ 0.67  1.09  0.8  66.72 66.72  8.98]
  [ 3.29  3.67  3.15 66.96 66.96  9.22]
  [ 5.83  6.16  5.38 65.98 65.98  8.24]]
```

```
[ [ 1.14  1.56  1.08 68.37 68.37 10.62]
  [ 0.74  1.11  0.84 66.73 66.73  8.98]
  [ 3.26  3.61  3.16 66.96 66.96  9.22]
  [ 5.87  6.17  5.42 65.98 65.98  8.24]]
```

```
[pattern_seed, day, sd_R] = [2, 7, 15]
```

```
max(u_0) = 168.8
O_threshold = 80
number of reward locations: 15
O_threshold = 90
number of reward locations: 12
O_threshold = 100
number of reward locations: 9
O_threshold = 110
number of reward locations: 6
target 1 in 4 DONE!
target 2 in 4 DONE!
target 3 in 4 DONE!
target 4 in 4 DONE!
```

```
-----
Value of Behaviour policy:57.758
```

```
O_threshold = 80
```

```
MC for this TARGET:[68.379, 0.181]
```

```
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.83, -1.32, -0.97]][[-68.38, -68.38, -10.62]
std:[[0.92, 0.97, 0.54]][[0.0, 0.0, 0.23]]
MSE:[[1.24, 1.64, 1.11]][[68.38, 68.38, 10.62]]
MSE(-DR):[[0.0, 0.4, -0.13]][[67.14, 67.14, 9.38]]
```

```
***
```

```
=====
```

```
O_threshold = 90
```

```
MC for this TARGET:[66.737, 0.191]
```

```
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-0.52, -0.95, -0.71]][[-66.74, -66.74, -8.98]
std:[[0.73, 0.72, 0.57]][[0.0, 0.0, 0.23]]
MSE:[[0.9, 1.19, 0.91]][[66.74, 66.74, 8.98]]
MSE(-DR):[[0.0, 0.29, 0.01]][[65.84, 65.84, 8.08]]
```

```
***
```

```
=====
```

```
O_threshold = 100
```

```
MC for this TARGET:[66.972, 0.196]
```

```
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-3.24, -3.57, -3.14]][[-66.97, -66.97, -9.21]
std:[[0.6, 0.6, 0.45]][[0.0, 0.0, 0.23]]
MSE:[[3.3, 3.62, 3.17]][[66.97, 66.97, 9.21]]
MSE(-DR):[[0.0, 0.32, -0.13]][[63.67, 63.67, 5.91]]
```

```
***
```

```
=====
```

```
O_threshold = 110
```

```
MC for this TARGET:[65.991, 0.195]
```

```
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
bias:[[-5.86, -6.16, -5.4]][[-65.99, -65.99, -8.23]]
std:[[0.81, 0.81, 0.57]][[0.0, 0.0, 0.23]]
MSE:[[5.92, 6.21, 5.43]][[65.99, 65.99, 8.23]]
MSE(-DR):[[0.0, 0.29, -0.49]][[60.07, 60.07, 2.31]]
```

```
***
```

```
=====
```

```
[ 1.04  1.51  1.1  68.36 68.36 10.63]
```

```
[pattern_seed, day, sd_R] = [2, 7, 20]
```

```
max(u_0) = 168.8
0_threshold = 80
number of reward locations: 15
0_threshold = 90
number of reward locations: 12
0_threshold = 100
number of reward locations: 9
0_threshold = 110
number of reward locations: 6
target 1 in 4 DONE!
target 2 in 4 DONE!
target 3 in 4 DONE!
target 4 in 4 DONE!
```

```
Value of Behaviour policy:57.768
```

```
0_threshold = 80
```

```
MC for this TARGET:[68.387, 0.236]
```

```
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
```

```
bias:[[-0.81, -1.29, -0.95]][[-68.39, -68.39, -10.62]]
```

```
std:[[1.13, 1.15, 0.71]][[0.0, 0.0, 0.27]]
```

```
MSE:[[1.39, 1.73, 1.19]][[68.39, 68.39, 10.62]]
```

```
MSE(-DR):[[0.0, 0.34, -0.2]][[67.0, 67.0, 9.23]]
```

```
***
```

```
=====
```

```
0_threshold = 90
```

```
MC for this TARGET:[66.745, 0.246]
```

```
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
```

```
bias:[[-0.5, -0.94, -0.7]][[-66.74, -66.74, -8.98]]
```

```
std:[[0.94, 0.9, 0.74]][[0.0, 0.0, 0.27]]
```

```
MSE:[[1.06, 1.3, 1.02]][[66.74, 66.74, 8.98]]
```

```
MSE(-DR):[[0.0, 0.24, -0.04]][[65.68, 65.68, 7.92]]
```

```
***
```

```
=====
```

```
0_threshold = 100
```

```
MC for this TARGET:[66.98, 0.25]
```

```
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
```

```
bias:[[-3.21, -3.55, -3.13]][[-66.98, -66.98, -9.21]]
```

```
std:[[0.77, 0.78, 0.57]][[0.0, 0.0, 0.27]]
```

```
MSE:[[3.3, 3.63, 3.18]][[66.98, 66.98, 9.21]]
```

```
MSE(-DR):[[0.0, 0.33, -0.12]][[63.68, 63.68, 5.91]]
```

```
***
```

```
=====
```

```
0_threshold = 110
```

```
MC for this TARGET:[65.999, 0.25]
```

```
[DR/QV/IS]; [DR_NO_MARL, DR_NO_MF, V_behav]
```

```
bias:[[-5.89, -6.17, -5.41]][[-66.0, -66.0, -8.23]]
```

```
std:[[1.05, 1.03, 0.73]][[0.0, 0.0, 0.27]]
```

```
MSE:[[5.98, 6.26, 5.46]][[66.0, 66.0, 8.23]]
```

```
MSE(-DR):[[0.0, 0.28, -0.52]][[60.02, 60.02, 2.25]]
```

```
***
```

```
=====
```

