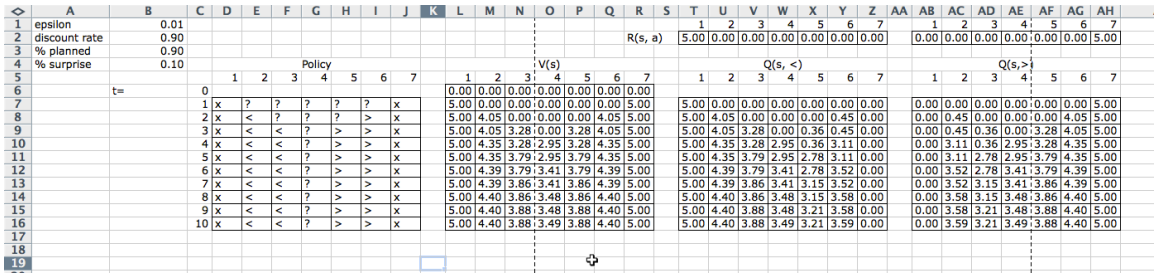
Module 4 – Self-Check

As noted in the video lectures, it is possible to solve value iteration for a 1-d Markov Decision Process in a spreadsheet by explicitly modeling the π, V, and Q arrays. You have been provided with a starter spreadsheet (module-4-self-check.xlsx)

For the self-check, you have been provided with a starter spreadsheet (module-4-self-check.xlsx). Fill in the formulas for Value Iteration (stochastic version) so that it looks like the following:



	A	B
1	epsilon	0.01
2	discount rate	0.90
3	% planned	0.90
4	% surprise	0.10

Policy

t=	1	2	3	4	5	6	7
0							
1	x	?	?	?	?	?	x
2	x	<	?	?	?	>	x
3	x	<	<	?	>	>	x
4	x	<	<	?	>	>	x
5	x	<	<	?	>	>	x
6	x	<	<	?	>	>	x
7	x	<	<	?	>	>	x
8	x	<	<	?	>	>	x
9	x	<	<	?	>	>	x
10	x	<	<	?	>	>	x

V(s)

1	2	3	4	5	6	7
0.00	0.00	0.00	0.00	0.00	0.00	0.00
5.00	0.00	0.00	0.00	0.00	0.00	5.00
5.00	4.05	0.00	0.00	0.00	4.05	5.00
5.00	4.05	3.28	0.00	3.28	4.05	5.00
5.00	4.35	3.28	2.95	3.28	4.35	5.00
5.00	4.35	3.79	2.95	3.79	4.35	5.00
5.00	4.39	3.79	3.41	3.79	4.39	5.00
5.00	4.39	3.86	3.41	3.86	4.39	5.00
5.00	4.40	3.86	3.48	3.86	4.40	5.00
5.00	4.40	3.88	3.48	3.88	4.40	5.00
5.00	4.40	3.88	3.49	3.88	4.40	5.00

R(s, a)

1	2	3	4	5	6	7
5.00	0.00	0.00	0.00	0.00	0.00	0.00

Q(s, <)

1	2	3	4	5	6	7
5.00	0.00	0.00	0.00	0.00	0.00	0.00
5.00	4.05	0.00	0.00	0.00	0.45	0.00
5.00	4.05	3.28	0.00	0.36	0.45	0.00
5.00	4.35	3.28	2.95	0.36	3.11	0.00
5.00	4.35	3.79	2.95	2.78	3.11	0.00
5.00	4.39	3.79	3.41	2.78	3.52	0.00
5.00	4.39	3.86	3.41	3.15	3.52	0.00
5.00	4.40	3.86	3.48	3.15	3.58	0.00
5.00	4.40	3.88	3.48	3.21	3.58	0.00
5.00	4.40	3.88	3.49	3.21	3.59	0.00

Q(s,>)

1	2	3	4	5	6	7
0.00	0.00	0.00	0.00	0.00	0.00	5.00
0.00	0.45	0.00	0.00	0.00	4.05	5.00
0.00	0.45	0.36	0.00	3.28	4.05	5.00
0.00	3.11	0.36	2.95	3.28	4.35	5.00
0.00	3.11	2.78	2.95	3.79	4.35	5.00
0.00	3.52	2.78	3.41	3.79	4.39	5.00
0.00	3.52	3.15	3.41	3.86	4.39	5.00
0.00	3.58	3.15	3.48	3.86	4.40	5.00
0.00	3.58	3.21	3.48	3.88	4.40	5.00
0.00	3.59	3.21	3.49	3.88	4.40	5.00

1. the cell B1 is named “discount_rate”, you can refer to it by that name in cell formulas, for example, “=discount_rate*10”
2. the cell B2 is named “planned”, you can refer to it by that name in formulas.
3. The cell B3 is named “surprise”, you can refer to it by that name in formulas.
4. if(condition,then,else) can be used to test values in cells. It can also be nested (the first one will need an =).
5. =max(*cellref1*, *cellref2*) will return the maximum value.
6. You do not need to fill in formulas for the goal states 1 or 7 except in the case of V(s).
7. Use “<” for “left”, “?” for “pick random” and “>” for “right” in your policy formula.
8. There isn’t a good way to control the number of iterations automatically, so epsilon is there to let you know when you should stop copying.

You should get the same results as above and converge in 10 iterations.