Microeconomic Foundations I:
Choice and Competitive Markets

# Student's Guide

# Appendix 6: Dynamic Programming

I present here the solutions to the problems left unsolved and/or given at the end of Appendix 6.

■ 3. ***The parking problem with uncertainty about the value of $\rho$.***

When there is uncertainty as to the value of $\rho$, the appropriate thing to do is to carry along as a state variable either (a) the posterior probability that $\rho = 0.7$ (or 0.9, one or the other), or (b) the number of spots so far observed that were occupied (or unoccupied). (The posterior probability is a sufficient statistic for everything you've seen so far, and the number of occupied spots so far observed [if you know where you are and when you began observing] allows you to compute the posterior.)

I'll work with the posterior: Let $C(s, p, n)$ be the expected cost (following the optimal policy) if you are at space $-n$, the spot is occupied ($s = 1$) or not ($s = 0$), and your posterior probability that $\rho = 0.7$ is $p$ (this assessed having seen the "condition" of the current spot). If the spot is empty and you park, you get a payoff of $n$. If either it is empty and you choose to go on to the next spot, or if it is full (so you must go on to the next spot), then the chance the next spot is occupied is

$$0.7p + 0.9(1 - p) = 0.9 - 0.2p,$$

if the next spot is occupied, the next posterior probability (by Bayes' rule) is

$$\frac{0.7p}{0.9 - 0.2p,}$$

and if the next spot is empty (with marginal probability $0.1 + 0.2p$), the next posterior probability is

$$\frac{0.3p}{0.1 + 0.2p.}$$

(Build a $2 \times 2$ probability table if you are unsure about the application of Bayes' Law here.) Hence Bellman's equation is

$$\begin{aligned} C(1,p,n) = &\quad (0.9-0.2p)\,C\left(1,\ \frac{0.7p}{0.9-0.2p},\ n-1\right) \\ &+\ (0.1+0.2p)\,C\left(0,\ \frac{0.3p}{0.1+0.2p},\ n-1\right), \text{and} \end{aligned}$$

$$\text{C(0,p,n)} = \min\{n, C(1,p,n)\}.$$

The boundary conditions are $C(0,p,0) = 0$ (if you are at spot zero and it is unoccupied, park), and $C(1,p,0) = p(10/3) + (1-p)(10/1)$. (To see the latter, note that if we are at spot 0, we will park at the first unoccupied spot we meet, so we are interested in the expectation of a random variable $X$, where $X$ is geometrically distributed (but is not 0) with unknown parameter $1 - d\rho$, and where $\rho = 0.7$ with probability $p$ and $= 0.9$ with probability $1-p$. To find the expectation of $X$, condition and then uncondition on the true value of $\rho$ to get the formula I've given.)

I doubt this problem can be solved analytically, so we have to resort to a numerical solution.

One way to try to solve this numerically is to solve for a discrete grid of posteriors, using linear interpolation (or something fancier) for finding the value function at posteriors that lie between values in your grid. The alternative, which I think will be more precise, is to note that, for the given prior 0.5 that $\rho = 0.7$ (prior to seeing the occupancy status of spot $-100$), there are two possible posteriors you can hold at spot $-100$ (two, because you have to condition on what you see at spot $-100$); at $n = -99$, three (not four!); at $n = -98$, four (not eight!); and so forth. The reason is that at $n = 97$ (say), there are only four possible results in terms of the posterior: All three spots were occupied, all three were empty, two were occupied, or one was occupied. The order of occupied and empty spots doesn't matter. So if you specify the prior, you can (almost surely) work your way through to a full solution, for the order of $n^2$ "states" that might occur.

This is how I tackled the problem, numerically. The first step was to build a table of posteriors. Rows correspond to the spot at which the posterior is being computed or, equivalently, the number of spots visited so far: Row 1 is spot $-100$, row 2 is spot $-99$, and so forth. Columns are the number of those spots that were occupied—for row 1 the columns are 0 and 1; for row 2, the columns are 0, 1, and 2; and so forth. Note that we expect the posterior that $\rho = 0.7$ to fall the more occupied spots we've seen. I won't depict the full table of posteriors, but Table GA6.1 shows how it begins.

initial prior: 0.5

TABLE OF POSTERIORS

| number of occupied spots / number of spots visited | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.750 | 0.438 | | | | | | | | | |
| 2 | 0.900 | 0.700 | 0.377 | | | | | | | | |
| 3 | 0.964 | 0.875 | 0.645 | 0.320 | | | | | | | |
| 4 | 0.988 | 0.955 | 0.845 | 0.585 | 0.268 | | | | | | |
| 5 | 0.996 | 0.984 | 0.942 | 0.809 | 0.523 | 0.222 | | | | | |
| 6 | 0.999 | 0.995 | 0.980 | 0.927 | 0.767 | 0.461 | 0.181 | | | | |
| 7 | 1.000 | 0.998 | 0.993 | 0.974 | 0.908 | 0.719 | 0.399 | 0.147 | | | |
| 8 | 1.000 | 0.999 | 0.998 | 0.991 | 0.967 | 0.885 | 0.666 | 0.341 | 0.118 | | |
| 9 | 1.000 | 1.000 | 0.999 | 0.997 | 0.989 | 0.958 | 0.857 | 0.608 | 0.287 | 0.094 | |
| 10 | 1.000 | 1.000 | 1.000 | 0.999 | 0.996 | 0.986 | 0.947 | 0.823 | 0.547 | 0.238 | 0.075 |

Table GA6.1. Posteriors for the Parking Problem

So, for instance, if you have visited 6 spots (have just seen spot $-95$ and 4 of those were occupied, you assess probability 0.767 that $\rho = 0.7$. But if 5 out of the 6 were occupied, your posterior assessment that $\rho = 0.7$ is 0.461.

With this table in place, we can proceed to compute the value functions. I found it easiest to have rows for the parking spot $(0, -1, -2, \ldots)$ and columns for the number of occupied spots observed up to but not including the current spot. Two tables were created of this sort, one for the case where the current spot is occupied, so the driver must drive on, and one for the case where the current spot is empty, so the driver has a choice to make. Since you never park before spot $-2$ when $\rho = 0.7$ or before spot $-6$ when $\rho = 0.9$, it is entirely intuitive that, in this case where $\rho$ is either 0.7 or 0.9, you never park before spot $-6$. What do you do? Consult Table GA6.2, which is a part of the "value if the spot is unoccupied" table of values.

Where the value in the table is the absolute value of the spot number, the optimal strategy is to park immediately, so the earliest you park is at spot $-6$, if you have (prior to this spot) seen 79 or more occupied spots. If you reach spot $-5$, you park first chance you get if you have seen 79 or more occupied spots. The same is true for spots $-4$ and $-3$. At spot $-2$, park if you can and if you have seen 78 or more occupied spots. At spot $-1$, park if you can, no matter how many occupied spots you have seen.

■ 7. ***The simplest multi-armed bandit problem.***

I will give the optimal strategy in terms of $p_t$, the Bayesian assessment made by the decision maker at date $t$ that the slot machine pays \$12 with probability 1/3. (Therefore, $p_0 = 0.8$.) Before doing this, let me say how $p_t$ evolves: On each turn, if you

| | | number of occupied spots observed prior to this | | | | |
|---|---|---|---|---|---|---|
| | | 77 | 78 | 79 | 80 | 81 |
| spot number | -10 | 6.521 | 6.554 | 6.563 | 6.565 | 6.566 |
| | -9 | 6.426 | 6.531 | 6.557 | 6.564 | 6.565 |
| | -8 | 6.137 | 6.462 | 6.539 | 6.559 | 6.564 |
| | -7 | 5.388 | 6.248 | 6.488 | 6.545 | 6.561 |
| | -6 | 4.125 | 5.642 | 6 | 6 | 6 |
| | -5 | 2.961 | 4.456 | 5 | 5 | 5 |
| | -4 | 2.333 | 3.186 | 4 | 4 | 4 |
| | -3 | 2.082 | 2.434 | 3 | 3 | 3 |
| | -2 | 1.984 | 2 | 2 | 2 | 2 |
| | -1 | 1 | 1 | 1 | 1 | 1 |
| | 0 | 0 | 0 | 0 | 0 | 0 |

Table GA6.2. Values at Unoccupied Spots as a Function of Occupied Spots Seen Prior to This. See the text for a fuller description, but this table implicitly identifies the optimal strategy.

choose not to play, you learn nothing, and so $p_{t+1} = p_t$. If you choose to play, the machine either pays \$0 or \$12. If it pays \$12, then you know that this is the sort of machine that will give \$12 with probability 1/3 each time; that is, $p_{t+1} = 1$, regardless of $p_t$ (as long as $p_t > 0$.) If it pays \$0, then the probability that it is the "good" type of machine becomes (by Bayes' rule)

$$p_{t+1} = \frac{(2/3)p_t}{(2/3)p_t + (1 - p_t)} = \frac{2p_t}{3 - p_t}.$$

Note that this posterior probability is monotonically increasing in the prior, is 1 if $p_t = 1$, is 0 if $p_t = 0$, and is strictly less than $p_t$ if $p_t$ is strictly betweeen 0 and 1. Moreover, if $p_t < 1$, then the ratio of $p_{t+1}$ to $p_t$ is $2/(3 - p_t)$, which is decreasing in $p_t$. Therefore, by a simple ratio test, we know that this sequence of posteriors decreases to zero, if the machine is played repeatedly and, each time, a \$0 reward is received.

With this as background, I assert the optimal strategy is to play the machine as long as $p_t > 1/13$, but to stop playing if and when $p_t \leq 1/13$.

I'll prove optimality of this strategy by proving that it is unimprovable. (Rewards are bounded above by \$11 per period and below by $\$ - 1$, and the discount factor is less than one.) I will do this in steps.

*Step 1. The (discounted, expected net present) value of following the strategy given if $p_t$ ever reaches 1 (because a \$12 prize is received) is \$30 (as a continuation value from that stage on).*

To see this, note that if $p_t$ reaches 1, it stays there forever. Therefore, the strategy says to play the game every time. The machine pays off an expected \$4 per round ((\$12)(1/3) + (\$0)(2/3)), netted against the \$1 it costs to play, so your expected net reward in each round is \$3. Discounted with discount rate 0.9, this has an discounted expected net present value of

$$\$3 + (0.9)\$3 + (0.9)^2\$3 + \ldots = \frac{\$3}{1 - 0.9} = \$30.$$

*Step 2. The (discounted expected net present continuation) value of following the strategy given is \$0 if* $p_t \leq 1/13$.

The strategy says not to play when $p_t \leq 1/13$. No information is received, so you will be choosing not to play next time and forevermore. So following this strategy nets \$0 each round, for a net present value of $0$.

*Step 3. The (discounted expected net present) value of following the strategy given is* $\geq$ \$0 *if* $p_t$ *is* $1/13$ *or more.*

This is the most complex step. Let me denote by $q_0$ the value $1/13$, and define inductively

$$q_{k+1} = \frac{3q_k}{2 + q_k}.$$

I assert that the sequence $\{q_k; k = 0, 1, \ldots\}$ has the properties that

(a) The sequence is increasing, with $q_k < 1$ for all $k$.

(b) The limit $\lim_k q_k = 1$.

(c) If $p_t = q_k$, the arm is pulled, and \$0 appears, then $p_{t+1} = q_{k-1}$.

In words, the $q_k$'s "invert" the Bayesian inference process described above. To prove parts a and b, consider the function $f(x) = 3x/(2+x)$, for $x \geq 0$. By inspection, $f(0) = 0$ and $f(1) = 1$. Also, $f'$ (the derivative of $f$) is $3/(2+x) - 3x/(2+x)^2 = 6/(2+x)^2 > 0$; $f$ is a strictly increasing function. Since $f(q_k) = q_{k+1}$, we know that the sequence of $q_k$'s is increasing, and since $f(1) = 1$, we know that if we start below 1, we do not exceed 1 in the sequence. That's part a. As for part b, suppose that the limit of the $q_k$'s was something less than 1, call it $q^*$. Then by the continuity of $f$, $f(q^*) = q^*$. But for any $q < 1$, $f(q)/q = 3/(2+q) > 1$; there are no fixed points of $f$ less than 1 (except for 0).

Part c is simple computation.

Now we proceed to prove step 3. Suppose $p_t > 1/13$. Suppose moreover that $p_t \in (q_0, q_1]$. Then the strategy calls for pulling the arm this turn and, if \$0 is the reward, since $p_t \leq q_0 = 1/13$, never playing again. The expected net (continuation) value is

$$(1/3)p_t(11 + (0.9)(30)) + (1 - (1/3)p_t)(-1 + (0.9)(0)), \qquad (*)$$

or the probability of getting a payback of \$12, which is $(1/3)p_t$, times the net immediate reward of \$11 and the discounted continuation value of \$30, plus the probability of getting back \$0, for a net of $-1$, times the discounted continuation value of \$0. Do the algebra, and you'll find that this is $13p_1 - 1$, which exceeds 0 as long as $p_t > 1/13$. Next suppose that $p_t \in (q_1, q_2]$. You do the same computations as above, except that in the equation $(*)$, the continuation value of 0 on the right-hand side is replaced by something greater or equal to 0, since $p_{t+1}$ will be in interval $(q_0, q_1]$. So this exceeds \$0 as long as $p_t > 1/13$. And so forth, "inductively," where the induction is on which interval $(q_k, q_{k+1}]$ contains $p_t$. (We know that $p_t$ is in some such interval because of part b above.)

*Step 4. The strategy is unimprovable (hence optimal).*

There are two cases: $p_t \le 1/13$ and $p_t > 1/13$.

Taking the former first, the strategy says not to play, for a next present value of \$0. The alternative is to play for one round. If \$12 is paid back, $p_{t+1} = 1$, so reverting to the strategy, you get a continuation value of \$30; if \$0 is paid back, $p_{t+1} < p_t < 1/13$, so reverting to the strategy says not to play, for a continuation value of \$0. Therefore, if you play this round, you net (in expectation)

$$(1/3)p_t(11 + 0.9 \cdot 30) + (1 - p_t/3)(-1 + 0.9(0)) = 13p_t - 1,$$

which is less than \$0 for $p_t < 1/13$ and just equal for $p_t = 1/13$. So for $p_t \le 1/13$, the strategy of not playing is unimprovable (in a single step).

And if $p_t > 1/13$, the strategy says to play. This, we showed in step 3, generates an expected discounted net present value in excess of \$0. (This is true as well for $p_t = 1$, although the appeal here is to step 1, not step 3.) If you do the alternative, which is not to play this round, you neither win nor lose anything this round and, since your posterior doesn't change, precisely beginning next round you get the expected discounted net present value from following the strategy today, but discounted by an additional 0.9. Hence instead of the continuation value $v$ from following the strategy, you get $0.9v$. Since $v \ge 0$, it is better (or, at least as good) to play today.

That does it.

■ 8. ***Another button-pushing, light-flashing machine.***

(a) Create two states, labeled $x$ and $y$, where the state at date $t$ is $x$ if the decision maker pushed $X$ at $t-1$, and $y$ if she pushed $Y$. Transition probabilities on the states are thus deterministic, and expected within-period rewards are $r(x, X) = 0.75(10) + 0.25(0) = 7.5$, $r(x, Y) = 12.5$, $r(y, X) = 2$, and $r(y, Y) = 7$, where (for instance $r(x, Y)$ is the expected reward if the state is $x$ and $Y$ is pushed).

I assert that if $\delta \ge 10/11$, the optimal strategy is to push $X$ regardless of state, and if $\delta \le 10/11$, the optimal strategy is to push $Y$ regardless of state. There is no claim here that these are the only optimal strategies, and in fact at $\delta = 10/11$, there are many others. But the problem asks you to produce an optimal strategy for each value of $\delta$, which is what I'm doing, once I verify optimality of these strategies. (I'm only interested in the cases where $\delta < 1$ here. If $\delta \geq 1$, any strategy gives you expected reward of $+\infty$.)

Regardless of $\delta$, the strategy of pushing $X$ every time gives expected total reward of $7.5{+}7.5\delta{+}7.5\delta^2{+}\ldots = 7.5/(1{-}\delta)$ starting in state $x$, and $2{+}7.5\delta{+}7.5\delta^2{+}\ldots = -5.5{+}7.5/(1{-}\delta)$ starting in state $y$. A one-step deviation in state $x$ gives $12.5 + 2\delta + 7.5\delta^2 + 7.5\delta^3 + \ldots$ starting in state $x$ and $7 + 2\delta + 7.5\delta^2 + 7.5\delta^3 + \ldots$ starting in state $y$, so this strategy is unimprovable if

$$7.5 + 7.5\delta \geq 12.5 + 2\delta \quad \text{and} \quad 2 + 7.5\delta \geq 7 + 2\delta,$$

both inequalities being true as long as $\delta \geq 5/5.5 = 10/11$.

(Now you fill in the details for $\delta \leq 10/11$ and the strategy of "always Y"; following the pattern set up in the preceding paragraph.)

(b) Now the states are called $xa$, if last time $X$ was pushed and the machine flashed $A$, and $z$, if last time $Y$ was pushed or the machine flashed $B$. Expected rewards are the same as in part (a), with $xa$ replacing $x$ and $z$ replacing $y$. But transition probabilities are different: From $xa$, if the decision maker pushes $Y$, then transition is to $z$ with certainty, and it is back to $xa$ with probability 0.75 (and to $z$ with probability 0.25) if she pushes $X$. From $z$, transition is to $z$ with certainty if the decision maker pushes $Y$, and to $xa$ with probability 0.2 if she pushes $X$.

I assert that "always push Y" is optimal, regardless of $\delta$.

First to find the value of this strategy. Starting in state $z$, this strategy gives an expected value of 7 each time, for a total (discounted) expected value of $7/(1-\delta)$. Starting in state $xa$, this strategy gives 12.5 in the first period and 7 thereafter, or $5.5 + 7/(1-\delta)$ in total.

I'll check unimprovability of this strategy starting in $xa$: If you deviate for one step, you get 7.5 in the first round and transition to $xa$ with probability 0.75 and to $z$ with probability 0.25, for a net expected value of

$$\begin{aligned} 7.5{+}\delta[(0.75)(5.5 + 7/(1-\delta)) + (0.25)7/(1-\delta)] &= 7/(1-\delta) + 0.5 + (0.75)(5.5)\delta \\ &= 7/(1-\delta) + 0.5 + 4.125\delta \leq 7/(1-\delta) + 4.625, \end{aligned}$$

which is always less than $5.5 + 7/(1-\delta)$. You still need to check unimprovability starting in $z$, but that is even easier.

■ 9. The optimal strategy is to pick $\beta$ whenever in state X. (There are no other choices to make.) Rewards in each period are bounded, and the discount factor (0.8) is strictly less than one, so the proof of optimality is a matter of checking that the strategy is unimprovable.

*Step 1. Compute the value of following the strategy of choosing $\beta$ whenever in state X.* If we follow this strategy, then at any point we are either in state X or state Y. I will let $v_X$ denote the expected value of all future rewards beginning with the current period if this period begins in state X, and I will let $v_Y$ denote the expected value of being in state Y. Then $v_X$ and $v_Y$ satisfy the following two recursive equations:

$$v_X = 2 + 0.8 v_Y \quad \text{and} \quad v_Y = 0 + 0.8(0.1 v_Y + 0.9 v_X).$$

To explain, the value starting in state X is an immediate payment of 2, and then next period you are in Y for sure, so you get the discounted value of being in state Y. And if you start in state Y, you get an immediate 0, and then, discounted by 0.8, the value of being in Y with probability 0.1 and the value of being in X with probability 0.9.

Now, the second of these equations is $v_Y = 0.08 v_Y + 0.72 v_X$ or $v_Y = (0.72/0.92) v_X$. And substituting this into the first equation gives

$$v_X = 2 + 0.8(0.72/0.92) v_X, \quad \text{or} \quad v_X = \frac{2}{1 - 0.8(0.72/0.92)} = 5.34883721,$$

and, therefore,

$$v_Y = \frac{0.72}{0.92} 5.34883721 = 4.18604651.$$

*Step 2. Check to ensure that this strategy is unimprovable (in a single step).* You only have a choice in state X, where the only alternative would be to choose $\alpha$. If, in state X, you choose $\alpha$ and then revert to the strategy I have hypothesized is optimal, you will net $1 + (0.8) v_X = 1 + (0.8)(5.34883721) = 5.27906977$, which is less than what you get by following the strategy. So the strategy in question is indeed unimprovable, and since rewards are bounded and discounted, its unimprovability implies that it is optimal.

■ 10. This is a discounted dynamic programming problem with an additive structure, where the discount factor 0.9 is strictly less than one, and where the per-period rewards are bounded: They are 0 if you haven't begun manufacturing, 2250 if you are manufacturing with a cost of 7, 9000 if you are manufacturing with a cost of 4, and 20250 if you are manufacturing with a cost of 1.

Therefore, all the tools of discounted dynamic programming are available. I use *unimprovable strategies are optimal*.

Moreover, the nature of the problem is very "transient." One you begin to manufacture the good, there are no further decisions to make. If you ever learn how to manufacture at a cost of 1, there is nothing more to learn, so it is evident that you should begin to manufacture immediately. If you ever learn how to manufacture at a cost of 4, then you never again need to be concerned about what happens if the cost is 7, since there are no circumstances in which that condition will ever recur.

I hypothesize that, if the cost of manufacture is 4 (and 1 is still possible), with $p$ your probability assessment that theoretically best cost (TBC) is 4 and $1-p$ that the TBC is 1, your optimal strategy is to continue to explore if $p \leq 37/45$ and to begin manufacture if $p > 37/45$. Let me describe how I found this number 37/45 and then prove my assertion.

Given $p$, if the strategy is to hold off for a round and then start manufacturing, there is probability $(1-p)/2$ that you will reduce costs to 1 and $p + (1-p)/2 = (1+p)/2$ that you will not. So your expected reward of following this plan is

$$0.9\left[\frac{1-p}{2}(20250 \times 10) + \frac{1+p}{2}(9000 \times 10)\right].$$

The 0.9 is the discount factor, while $20250 \times 10$ is the net present value (next period) of manufacturing with a cost of 1, and $9000 \times 10$ is the NPV (next period) of manufacturing with a cost of 4. The alternative is to begin manufacturing immediately, with an NPV (this period) of 90,000. So I find the implications for $p$ of the first quantity being greater or equal to the second, which is

$$.9\left[\frac{1-p}{2}(20250 \times 10) + \frac{1+p}{2}(9000 \times 10)\right] \geq 90000,$$

and this gives $p \leq 37/45$.

Note that this calculation verifies the unimprovability of my announced strategy for $p \geq 37/45$, since when I follow this strategy, my value function for $p \geq 37/45$ is 90,000 (the strategy says, start to manufacture immediately), and I've shown that for $p$ in this range, beginning to manufacture is better than the one-step alternative of exploring for a round and then (since, if I don't move costs down to 1, my new assessment will be greater than my previous $p$, hence greater than 37/45) beginning to manufacture.

But it doesn't verify unimprovability for $p \leq 37/45$. To do that, I need the value of following this policy or, rather, I need to know that the value is greater than 90,000, which is the value if, instead of exploring, I move to the (irreversible) strategy of beginning to manufacture. I can show this by an argument of the sort used in the "simplest multi-armed bandit problem." But, in this case, I can revert to a more computational approach:

If ever costs fall from 7 to 4, at that point, $p = 0.5$; the chances of improving costs from 7 to 4 are independent of whether the TBC is 4 or 1, so their relative likelihoods stay the same (equal), and so the moment you learn that TBC is not 7, you are back to a 50-50 assessment on whether it is 4 or 1. Now if you explore for one round, either costs fall to 1 or they don't, and a simple application of Bayes' rule tells you that, if they don't, $p$ rises to $2/3$. Another round of unsuccessful exploration leads to $p = 4/5$. And successive rounds give $p = 8/9$, $p = 16/17$, $p = 32/33$, and so forth. Why? The general rule is if $p = 2^n/(2^n + 1)$ in this round, then the chance of failure (no lowering of cost) this round is $2^n/(2^n + 1) + (1/2)(1/(2^n + 1)) = (2^{n+1} + 1)/(2^{n+1} + 2)$ and the posterior probability (given this event) that TBC is 4 is

$$\frac{2^n/(2^n+1)}{(2^{n+1}+1)/(2^{n+1}+2)} = \frac{2^{n+1}}{2^{n+1}+1}.$$

At $p = 8/9$, we are in the region where $p > 37/45 = 0.8222$, so we know that, in following my hypothesized strategy, we search for a lower cost (of 1) immediately, when $p = 2/3$, and when $p = 4/5$, but abandon the search and start manufacturing if we fail to lower cost on the third try (after costs are lowered to 4).

We can then use a simple finite-horizon recursion to compute the value of following this strategy. When $p = 4/5$, we get

$$.9\left[\frac{1}{10}202500 + \frac{9}{10}90000\right] = 91125.$$

This then gives us the value for when $p = 2/3$, as

$$.9\left[\frac{1}{6}202500 + \frac{5}{6}91125\right] = 98718.75$$

and, for when $p = 1/2$,

$$.9\left[\frac{1}{4}202500 + \frac{3}{4}98718.75\right] = 112197.656.$$

Now we can go back to the situation where best current cost is 7, and TBC could be 7, 4, or 1. Now, if a "breakthrough" happens and costs reduce to 4, the continuation value is 112197.656, so you can repeat the analysis given above to discover that you should keep trying to lower costs as long as $q$, the current posterior assessment that 7 is the TBC, exceeds 0.94425718. (That is, 0.94425718 is the solution to the equation $22500 = .9[22500(1 + q^*)/2 + 112197.656(1 - q^*)/2.)$

Given a starting value of $q = 1/3$ (with the assessments that TBC = 4 and that TBC = 1 always half of $1 - q$, as long as current cost is 7), the sequence of posteriors for $q$ as long as no "breakthough" to a cost of 4 occur, is 1/2, then 2/3, then 4/5, 8/9, 16/17 = 0.94117697, and 32/33 = 0.96969696.... So, by the same logic as before, you will keep trying to lower costs for six periods, giving up (and starting to manufacture) when $q$ reaches 32/33. To verify unimprovability of this, we need the values for each posterior, which are computed in the same fashion as before:

$$\text{for } q = \frac{16}{17}, \text{ the value is } .9\left[\left(1 - \frac{1}{34}\right)22500 + \frac{1}{34}112197.656\right] = 22624.3497,$$

$$\text{for } q = \frac{8}{9}, \text{ the value is } .9\left[\left(1 - \frac{1}{18}\right)22624.3497 + \frac{1}{18}112197.656\right] = 24849.5801,$$

and, similarly,

for $q = 4/5$, the value is 30218.6589,

for $q = 2/3$, the value is 39493,6426,

for $q = 1/2$, the value is 51902.6814, and

for $q = 1/3$, the value is 64800.9057.

Note that these computations carry with them the test of unimprovibility: For $q = \frac{1}{3}, \frac{1}{2}.\frac{4}{5}.\frac{8}{9}$, and $\frac{16}{17}$, the strategy is to explore, and we've computed the values from following this strategy (together with the "second half" of it, done before) as all being greater than 22500, the value of the alternative (manufacture immediately). While for any $q > 0.94425718$, which includes $q = \frac{32}{33}$, the inequality we implicitly computed previously shows that manufacturing is better than a one-step change of explore and then, if you fail to lower costs to 4, manufacture.

**■ 11. Forest Management.**

(a) The optimal strategy is to cut the forest as soon as it reaches size 19 (or more).

If you following this strategy, the forest will be of size 19 ($= 1 + 2 + 4 \times 4$) each time it is cut, and you will have a profit of 18. The value of following this strategy can be written as $v_n$, the value of having a forest of size $n$ going into a period (after the growth has taken place), for $n = 1, 3, 7, 11, \ldots$. Then

$$v_1 = .9^5(18 + .9^6 \times 18 + .9^{12} \times 18 + \ldots) = \frac{.9^5 \times 18}{1 - .9^6} = 22.684059,$$

$$v_3 = .9^4(18 + .9^6 \times 18 + \ldots) = 25.20451,$$

and similarly

$$v_n = \begin{cases} 28.005, & \text{for } n = 7, \\ 31.116679, & \text{for } n = 11, \\ 34.5740878, & \text{for } n = 15, \\ 38.415631, & \text{for } n = 19. \end{cases}$$

For larger $n$ (of the form $n = 3 + 4m$ for some integer $m > 4$), this strategy says to cut immediately, for a value of

$$v_n = n - 1 + .9^6 \times 18 + .9^{12} \times 18 + \ldots = n + 19.4156531.$$

To show the optimality of this strategy, I must show that it is unimprovable. For a forest of size 1, 3, 11, or 15, the strategy calls for letting the forest grow, while a (one-step) variation would call for cutting the forest immediately. Cutting the forest immediately generates $n - 1$ immediately (where $n$ is the size of the forest), and then (reverting back to the strategy) 18 in six periods, 18 again in twelve, etc., for a net present value of $19.4156531 + n$ (gotten the same way was we obtained $v_n$ for $n > 19$). That is, the variant strategy nets

$$v'_n = \begin{cases} 19.4156531, & \text{for } n = 1, \\ 22.4156531, & \text{for } n = 3, \\ 26.4156531, & \text{for } n = 7, \\ 30.4156531, & \text{for } n = 11, \text{ and} \\ 34.4156531, & \text{for } n = 15. \end{cases}$$

Therefore, none of these variations is a one-step improvement on the allegedly optimal strategy. As for forests of size $n \geq 19$, the strategy says to cut them, so the one-step variation is to let them go for a year and then cut. This nets $.9(19.4156531 + n + 4) = 21.0740679 + .9n$, versus $19.4156531 + n$ which is obtained by following the allegedly optimal strategy. The former is less than the latter as long as $1.6584148 \leq .1n$ or $n \geq 16.584148$, which is true for these $n$.

This shows that the announced strategy is unimprovable, and therefore it is optimal.

(b) I assert that the optimal strategy is to wait until the forest reaches size 15 (or more), and then cut it on the first opportunity where $p_t = 1.5$. That is, you never cut the forest if $p_t = 1$, and you never cut it before it reaches size 15.

To show that this strategy is optimal, I first have to find out the value of following it. Begin with a forest of size $n \geq 15$. I'm going to evaluate the expected value of following the strategy, unconditional on the current price, calling this $v_n$. The strategy says to cut the forest on the first opportunity when $p_t = 1.5$, which means that you will cut the forest immediately with probability 1/2, in one period with probability 1/4, and so forth. Therefore, your expected net present value is

$$v_n = \frac{1}{2}\left[1.5(n-1)+.9^5 v_{15}\right]+\frac{.9}{4}\left[1.5(n+4-1)+.9^5 v_{15}\right]+\ldots+\frac{.9^k}{2^{k+1}}\left[1.5(n+4k-1)+.9^5 v_{15}\right]+\ldots.$$

To explain, with probability 1/2, the immediate price is 1.5 and you cut the forest down, netting $1.5(n-1)$ immediately and, beginning in five periods, $v_{15}$; with probability 1/4 the immediate price is 1 but the next price is 1.5, so you cut next time, netting an immediate $1.5(n+4-1)$ and, with a further five period delay, $v_{15}$, and so forth. Replacing $n$ with 15 in the formula gives a recursive equation for $v_{15}$, namely

$$v_{15} = \frac{1}{2}\left[1.5(14)+.9^5 v_{15}\right]+\frac{.9}{4}\left[1.5(14+4)+.9^5 v_{15}\right]+\ldots+\frac{.9^k}{2^{k+1}}\left[1.5(14+4k)+.9^5 v_{15}\right]+\ldots,$$

which is

$$v_{15} = 21\times\left(\frac{1}{2}+\frac{.9}{4}+\ldots\right)+\left(\frac{.9}{4}\times 6+\frac{.9^2}{8}\times 12+\ldots\right)+.9^5 v_{15}\left(\frac{1}{2}+\frac{.9}{4}+\ldots\right).$$

This is

$$v_{15} = \frac{10.5}{1-.45}+\frac{3\times .45}{.55^2}+\frac{.9^5 v_{15}}{2}\frac{1}{1-.45} = 19.090909+4.46281+.53680909 v_{15}.$$

Therefore,

$$v_{15} = \frac{19.090909+4.46281}{1-.53680909} = 50.85099588.$$

This then allows us to compute $v_n$ for $n > 15$:

$$v_n = (1.5(n-1)+.9^5 v_{15})\left(\frac{1}{2}+\frac{.9}{4}+\ldots\right)+\left(\frac{.9}{4}\times 6+\frac{.9^2}{8}\times 12+\ldots\right) =$$

$$\frac{.75(n-1)+.9^5\times 50.85099588\times .5}{.55}+4.46281 = 1.3636n+30.39645,$$

which gives

$$v_{19} = 56.30485, v_{23} = 61.75925, v_{27} = 67.21365,$$

and so forth. For $n < 15$, the value of following the strategy is found by

$$v_n = \begin{cases} 0.9v_{15} = 45.7659, & \text{for } n = 11, \\ 0.9^2 v_{15} = 41.1893, & \text{for } n = 7, \\ 0.9^3 v_{15} = 37.07038, & \text{for } n = 3, \text{ and} \\ 0.9^4 v_{15} = 33.3633384, & \text{for } n = 1, \end{cases}$$

since in each case the strategy calls for waiting until the forest grows to size 15.

Now to verify unimprovability. First we'll check for sizes below 15. The strategy says to wait, regardless of the current price of wood, and so if cutting when the price is 1.5 is not a one step improvement, then neither is cutting when the price is 1. When the forest is of size 1, if the price is 1.5, cutting nets 0 immediately and $v_{15}$ in five periods, which is clearly worse than getting $v_{15}$ in four periods, which is what the strategy causes. So cutting when the forest is of size 1 is not a one-step improvement.

When the forest is of size 3 and the price is 1.5, cutting nets an immediate 3 plus $v_{15}$ in five periods, or $3 + .9^5 v_{15} = 33.0270046$, which is worse than following the allegedly optimal strategy. Similar calculations verify unimprovability of the strategy when $n = 7$ and 11.

When the forest is of size 15, the strategy calls for cutting if the current price is 1.5 and leaving the forest if the current price is 1. So if the current price is 1.5, not cutting gives $.9v_{19}$ next period, which is $.9 \times 56.3055413 = 50.674365$, which is worse than carrying out the strategy, netting an immediate 21 plus $.9^5 v_{15}$, for a total of 51.027.

And when the forest is of size $n \geq 15$, the strategy calls for cutting if the current price is 1.5 and letting the forest grow for another period if the current price is 1. Suppose the current price is 1.5. Then following the strategy nets $1.5(n-1)$ immediately and a continuation value of $.9^5 v_{15}$, for a total $28.52700456 + 1.5n$. Letting the forest grow for another period nets an expected $.9v_{n+4} = .9(1.3636(n+4) + 30.39645) = 1.22724n + 32.265765$. So following the strategy is better. And if the price is 1, cutting the forest immediately nets $n - 1$ immediately and $.9^5 v_{15}$ as a continuation value, or a total of $n + 29.02668$, versus $1.22724n + 32.265765$ by following the strategy (of waiting until next period). Following the strategy is clearly better.

(c) I assert that the optimal strategy is to cut the forest either the first time it fails to grow or when it reaches size $n \geq 15$.

First step in verifying this is to compute the value function for following this strategy. Let $v_n$ be the value beginning with a forest of size $n$ that has just grown, let $v'_n$ be the value of a forest of size $n$ that has not grown, and let $v$ be the continuation value following this strategy, immediately after cutting and replanting the forest.

Now consider the continuation value $v$ obtained immediately after cutting the forest. The forest grows to sizes 1 and 3 in the next two periods. Then, in the third period, it doesn't grow at all with probability 0.2, and you harvest it and replant. You harvest and replant (at the size 7) in period 4 with probability $0.8 \times 0.2 = 0.16$; you harvest and replant at the size 11 in period 5 with probability $0.8^2 \times 0.2 = .128$; and you harvest and replant a forest of size 15 in period 5 with probability $0.8^3 = 0.512$. Therefore,

$$v = (0.9^3)(0.2)(2 + v) + (0.9^4)(0.16)(6 + v) + (0.9^5)(0.128)(10 + v) + (0.9^5)(0.512)(14 + v).$$

Collecting terms this is

$$v = 5.90991552 + 0.6286896v, \quad \text{or} \quad v = \frac{5.90991552}{1 - 0.6286896} = 15.9163749.$$

For $n \geq 15$, the strategy calls for cutting the forest whether it has just gown or not, so $v_n = v'_n = n - 1 + 15.9163749 = n + 14.9163749$. For any $n$, if the forest has not grown, the strategy says to cut immediately, so $v'_n = n + 14.9163749$. The harder values to compute are $v_n$ for $n < 15$. In all such cases, the strategy says not to cut. Therefore:

- For $n = 11$, $v_{11} = 0.9(0.8v_{15} + 0.2(10+v)) = 0.9(0.8 \times (14+v) + 0.2 \times (10+v)) = 26.72306488$.
- For $n = 7$, $v_7 = 0.9(0.8v_{11} + 0.2(6 + v)) = 23.18555419$.
- For $n = 3$, the forest is certain to grow to size 7, so $v_3 = 0.9v_7 = 20.86699877$.
- For $n = 1$, the forest is certain to grow to size 3, so $v_1 = 0.9v_3 = 18.78029890$.

Now to verify unimprovability: For a forest of size 1, the alternative to the allegedly optimal strategy is to cut immediately, yielding 0 immediately and a continuation value of 15.9163749. Waiting is better.

For a forest of size 3, the alternative strategy, cutting immediately, nets 2 immediately and the continuation value of 15.9163749, for a total of 17.9163749, clearly worse than following the strategy.

For any forest of size $n > 3$ that has stopped growing, the strategy says to cut immediately. The alternative is not to cut, which simply delays by one period the immediate value and the continuation, since the forest is certain not to grow. So it is clearly the case that the alternative of not cutting for one stage is inferior to immediately cutting.

For a forest of size 7 that has just grown, cutting immediately nets $6 + 15.9163749$, versus 23.18... for following the strategy. Following the strategy is better.

For a forest of size 11 that has just grown, cutting immediately nets $10 + 15.9163749 = 25.9163749$, versus 26.73... for following the strategy. Following the strategy is better.

For a forest of size $n \geq 15$ that has just grown, the strategy says to cut immediately, netting $n + 14.9163749$. The alternative one-step deviation is to wait a period (following which you will cut immediately, since you revert to the strategy): This nets

$$0.9\big(0.8(n+3+15.9163749)+0.2(n-1+15.9163749)\big) = 0.9(n+18.1163749) = 0.9n+16.3047374.$$

Following the strategy is at least as good as long as

$$0.1n \geq 16.3047374 - 14.9163749 = 1.38836251 \quad \text{or} \quad n \geq 13.8836251,$$

which is true.

(d) In this case, you have to keep track of how many times in a row the forest has experienced zero growth and the resulting posterior probability that it is in no-growth mode. Because a growth of 4 is incompatible with no-growth phase, every time you see growth of 4, you reassess that, for that period, the forest was indeed in growth mode, and so the probability it is in growth mode in the next period is 0.8, and the probability of a growth of 4 units is 0.72. But suppose the forest, in this next period, didn't grow (grew 0 units). The posterior probability that it was indeed in growth mode is

$$\frac{\text{Prob(0 units and growth mode)}}{\text{Prob(0 units)}} = \frac{0.8 \times 0.1}{0.8 \times 0.1 + 0.2 \times 1} = \frac{0.08}{0.28} = 0.2857,$$

and hence the probability that it is in growth mode next period is $.8 \times 0.2857 = 0.22857$, and the probability of 4-unit growth in that period is $0.22857 \times 0.9 = 0.2057$. If again it doesn't grow, the posterior probability that it was in growth mode, computed by the same Bayes' rule calculations, falls to 0.02877698, so the probability (following this) that it is in growth mode the next period is 0.02302158 and the probability that it grows 4 units is 0.020719. And so forth.

My initial hypothesis as to the optimal strategy is: As long as the forest keeps growing, let it grow to size 15 and then cut. If it ever fails to grow, cut immediately. Actually, my strategy needs to be specified a bit better than that: Precisely, it is to cut any forest that is of size 15 or larger, and cut any forest that has failed to grow in the last period, but let a forest of size 11 or less go if it has just grown. You can see this strategy depicted in Figure GA6.1, where I have drawn an "event tree" out to period 6 after a cutting. The open nodes represent states where the strategy says to let the forest grow, and the closed nodes are states where the strategy says to cut.

If I follow this strategy, I can evaluate the continuation value just after a cutting as the solution $v$ to the following equation:

$$\begin{aligned} v =& (.72)(.72)(.72)(.9^5)(14+v) + (.72)(.72)(.28)(.9^5)(10+v) \\ & (.72)(.28)(.9^4)(6+v) + (.28)(.9^3)(2+v). \end{aligned}$$

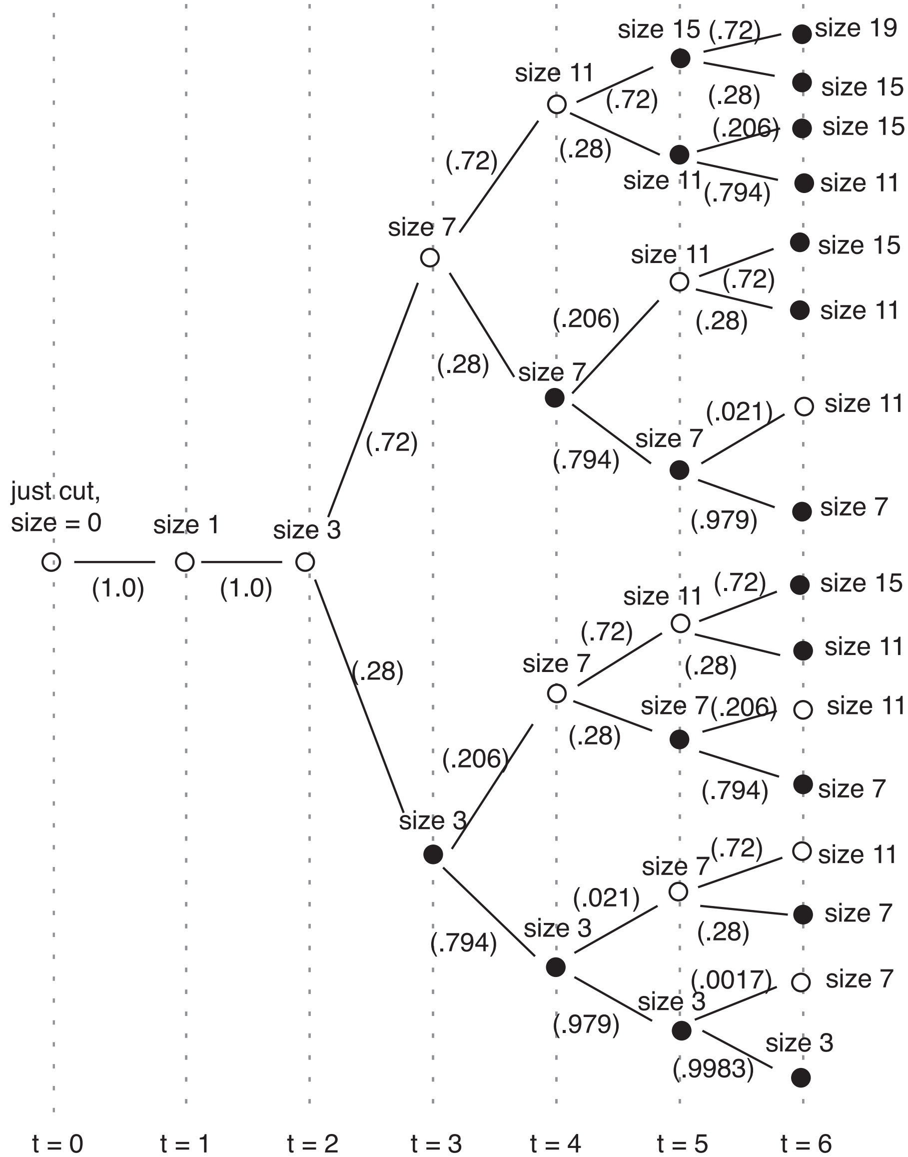


Figure GA6.1. The Optimal Strategy for Harvesting the Forest in Part d. An open node means to let the forest continue to grow, while a filled-in node means a position in which the forest is cut. Transition probabilities are indidcated on branches in parentheses.

This is just evaluating each branch that might occur, following this strategy. This equation, if you do the math, becomes

$$v = 5.14455557 + .64249978v \quad \text{which gives} \quad v = 14.3903562.$$

To evaluate one-step deviations from this strategy, we have to know the value of the hypothesized strategy from positions where you let the forest grow, according to the strategy. This includes situations where the forest has just grown and is of size 1, 3, 7, and 11.

- If the forest has just grown (so you know it is in growth mode) and is of size 11, you are supposed to leave it. It grows again with probability 0.72, and you get $14 + v$, and it fails to grow with probability .28, you cut it, and you net $10 + v$. So the value to you of being in this position is

$$.9\left[(.72)(14 + v) + (.28)(10 + v)\right] = 24.5433206.$$

- If the forest has just grown and is of size 7, you leave it. It grows again with probability 0.72, giving you a full continuation value of 24.5433206. It fails to grow with probability 0.28, so you cut it, netting $6 + v$. So the full value to you of being in this position is

$$.9\left[(.72)(24.5433206 + (.28)(6 + v)\right] = 21.0424415.$$

- If the forest is of size 3, you leave it. It grows to size 7 with probability .72, netting you the value just computed. It remains at size 3, in which case you are meant to cut it, netting $2 + v$. So your net value is

$$.9\left[(.72)(21.0424415) + (.28)(2 + v)\right] = 17.7658719.$$

- If the forest is of size 1, you leave it. It is sure to be size 3 next time, so your value in this position is $.9 \times 17.7658719 = 15.9892847$.

Now we can verify unimprovability. First we'll verify that, when the strategy says to let the forest go for another period, it isn't better to cut immediately. This involves the four situations just evaluated.

- If the forest is of size 11, having just grown, and you cut, you net $10+v = 24.3903562$, which is just slightly worse than the 24.5433206 you get by following the strategy.
- If the forest is of size 7, having just grown, and you cut, you net $6 + v = 20.3903562$, which is worse than following the strategy and getting 21.0424415.
- If the forest is of size 3, having just grown, cutting nets $2 + v = 16.3903562$. Following the strategy is better.
- If the forest is of size 1, cutting nets $v = 14.3903462$. Following the strategy is better.

Now for forests of all sizes 15 or more. The strategy says to cut immediately, which nets $n - 1 + v = 13.3903562 + n$, where $n$ is the size of the forest. The alternative is to let the forest go a period, and at best it will grow by 4 units with probability 0.72 and not at all with probability 0.28, after which (reverting to the strategy) you cut. So the alternative at best yields

$$.9\big[(.72)(n + 17.3903562) + (.28)(n + 13.3903562)\big] = .9n + 14.6433206.$$

Following the strategy is at least as good as long as $.1n \geq 1.25296438$, or $n \geq 12.529$, which is certainly true.

For the smaller sized forests, the question is whether it is better to cut or to leave the forest after it has failed to grow. This only affects forests of size 11, 7, and 3.

- For a forest of size 11 that has just failed to grow, at best the odds of it growing next period is 0.2057. If you let it grow and then revert to strategy, you will cut it next period whether it grows or not. So letting it grow yields

$$.9\big[(0.2057)(14 + v) + (0.7943)(10 + v)\big] = 22.6918406,$$

  while cutting it immediately nets $10 + 14.3903562$; cutting it immediately is better.

- For a forest of size 7 that has just failed to grow, the most optimistic probability of growth in the coming period is 0.2057, so at most the value of letting it grow is

$$(.9)\big[(0.2057)(24.5433206) + (0.7943)(6 + v)\big] = 19.1201589,$$

  versus the 20.3903562 that you get from cutting it, so cutting it is better.

- And for a forest of size 3 that has just failed to grow, the most optimistic value of leaving it for a year (and reverting to the strategy) is

$$(.9)\big[(0.2057)(21.0424415) + (0.7943)(2 + v)\big] = 15.6125611,$$

  versus the 16.3903562 that you get from cutting it, so cutting it is better.