Dynamic Choice Over Menus

A companion to Francetich and Kreps (2013)

Alejandro Francetich *†

Department of Decision Sciences and IGIER, Bocconi University, Italy

April 9, 2014

Abstract

A decision maker can choose *up to* two alternatives, or "tools," over time. The rewards from these choices depend on an unobserved state of nature. There are two possible states, and one and only one tool is profitable in each state. Opportunities to "employ" or draw value from the favored tool obey a Poisson process with known arrival rate, but the identity of the favored tool is unobserved. The decision maker only observes the realized rewards of the tools chosen, and choosing each tool entails a "rental" cost. The problem is a multi-armed bandit problem, where the arms are the possible subsets of tools. These arms are not independent: Choosing both tools simultaneously provides information about each individual tool. Applications include hiring of experts by professional-services firms.

Keywords: Two-armed bandits, experimentation, exploitation, correlated arms

JEL Classification Numbers: D83, D90

*Via Röntgen 1, 20136 Milan, Italy. Email address: alejandro.francetich@unibocconi.it

†This paper is a companion to Francetich and Kreps (2013), and borrows from Chapter 3 of my dissertation, submitted to the Stanford Graduate School of Business. I am deeply indebted to David Kreps for his guidance. I also thank Sven Rady for his encouragement and support. I gratefully acknowledge financial support from ERC advanced grant 324219. Any remaining errors and omissions are all mine.

1 Introduction

Imagine a hiker who has recently moved to a cabin near the woods, and who goes out hiking every day. She faces challenges while out in the wilderness, challenges she can overcome using tools. These tools are to be rented for the day from a local hardware store, so it is costly to be "well equipped"; alternatively, if the tools are owned by the decision maker, they may be too heavy to carry them all at once all the time. Once faced with the current challenge, the hiker can employ the tool that best serves the day's challenge *out of those that she has in her possession*. However, she is uncertain about how useful a tool is on any given day. Her problem is to select, every morning, which tools to rent for her walk. By the end of each day, she observes the value of the tools she carried, and can use this information to guide future rental decisions.

Alternatively (and perhaps less fancifully), consider the problem of a professional-services firm, such as a consultancy or a legal partnership, employing a pool of experts to serve their clients. Imagine that only one case is handled on each day, and that this case can only be dealt with by a single staff member. The manager, after an initial interview with the client, learns the specifics of the case and assigns it to the member of her staff who is best suited for the specific job. Of course, only experts who are on staff are available for this specific job; hiring new experts takes time. But keeping a large staff, one that can handle all conceivable cases, is costly: Experts that remain idle during any period must nonetheless be paid a wage.

These problems share the following basic structure. A decision maker (hiker, firm manager) faces a set of alternatives (tools, experts), and chooses subsets or "menus" (tool bundles, teams of experts) from this set. Opportunities to "employ" or draw value from the chosen menu (hiking hazards, firm clients) obey an unknown stochastic process. The decision maker only observes the realizations of these process corresponding to the alternatives in the chosen menu (tools/experts performance), and choosing larger menus is costly (in terms of rental cost or wages).

Thus, there is a trade-off between gathering more information by choosing larger menus, on the one hand, and saving on cost by choosing smaller menus, on the other hand. This "exploration-exploitation" trade-off makes the problem a multi-armed bandit problem, where the arms are the possible menus.[1] However, these "arms" are not independent: Rewards from overlapping menus are correlated, even if the rewards from individual elements are independent.

[1]Alternatively, if we identify each tool with an arm, we can think of the problem as a multi-choice multi-armed bandit problem; the decision maker faces a finite set of arms and may choose several of them at a time. On multi-choice multi-armed bandits, see Bergemann and Valimaki (2001).

This paper analyzes the following continuous-time instance of this formal structure. There are two tools and two possible states of nature. Opportunities to employ these tools arrive over time according to a known Poisson process, but correspond to each of the tools according to the unobserved state of nature. Over any time interval, the decision maker can only observe the arrival of opportunities for tools selected. Therefore, by choosing a single tool, she cannot distinguish between the arrival of an opportunity for the other tool from failure of arrival altogether.

Of course, while she may only observe the realization for a single tool, the decision maker can still make inferences about the distribution of values for the other tool due to their correlation. In general, these forms of dependence can preclude a complete characterization of optimal strategies in multi-armed bandit problems. In fact, the problem of characterizing optimal strategies in general dependent multi-armed-bandit problems is an open problem. Therefore, beyond the relatively simple structure of the problem in this paper, the traditional approach to analyzing such problems is extremely limited. Thus, in Francetich and Kreps (2013), we turn to *heuristics*: simple rules of behavior featuring self-correcting techniques based on past experience to improve future performance.

While the spirit of the problem studied in this paper is closely related to Francetich and Kreps (2013), the formal techniques employed to analyze the problem, as well as the mathematical structure of the solutions, borrow heavily from Keller and Rady (2010) and Klein and Rady (2011).

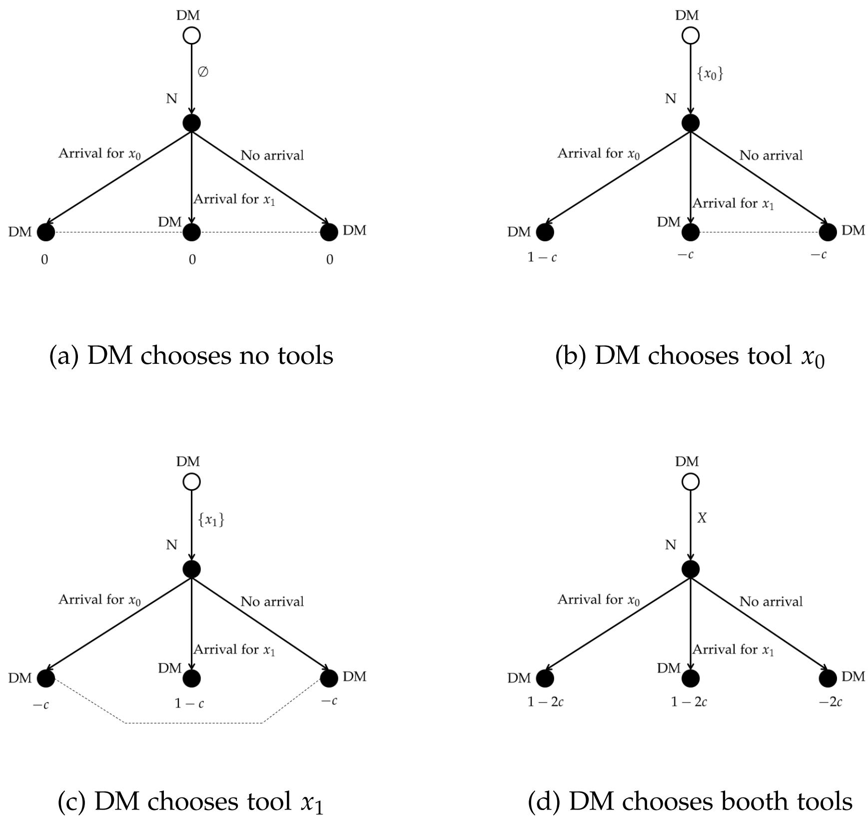
The rest of the paper is organized as follows. Section 2 describes the formal framework of the problem. Section 3 describes Bayesian updating and presents the Bellman equation for the problem. Section 4 describes the solution to the Bellman equation. Finally, section 5 concludes. Proofs are relegated to the appendix.

2 Framework

There is a set of alternatives $X = \{x_0, x_1\}$, representing "tools" a decision maker (henceforth, DM) can employ. DM allocates her time between the different subsets of X. The set of allocations of a unit of time between the bundles $\{x_0\}, \{x_1\}$, and $\{x_0, x_1\}$ is denoted by $A := \{\alpha \in [0,1]^3 : \alpha_1 + \alpha_2 + \alpha_3 \leq 1\}$ — if $\alpha \in A$, the residual $1 - \alpha_1 - \alpha_2 - \alpha_3$ is the fraction of time spent on the empty set.

Tools must be rented to be employed; $c > 0$ is the per-tool rental rate. A tool employed yields a gross reward of 1. Opportunities to employ these tools arrive as follows. There are two possible states of nature $\omega \in \{0, 1\}$, the realization of which is unobserved by DM. Employment opportunities for tool x_0, x_1 arrive according to Poisson processes with arrival rates $\lambda^0\omega$ and $\lambda^0(1-\omega)$, respectively, where $\lambda^0 > 0$ is a *known* arrival rate. Figure 1 describes the timing and information structure of the problem.

A more flexible specification would allow for $\omega \in (0,1)$. Under this alternative specification, opportunities can arrive for both tools. We can think of this alternative process as having nature first drawing an opportunity from a Poisson process with arrival rate λ^0, and the "allocating" this opportunity to tool x_0 or x_1 with probabilities $\omega, 1-\omega$, respectively, independently of past arrivals and allocations. Thus, we get a partitioning of the Poisson process. However, this additional flexibility comes at the cost of slowing down the learning process without providing significant new insights. In this alternative scenario, the arrival of an opportunity for a tool is no longer indicative of the tool being the superior one, and the information previously conveyed by a single success on a tool is now conveyed by a larger frequency of arrivals for said tool relative to the other tool over a long period of time.



(a) DM chooses no tools

(b) DM chooses tool x_0

(c) DM chooses tool x_1

(d) DM chooses booth tools

Figure 1: DM's observations and payoffs under each of her possible choices. The dotted lines represent information sets, namely, nodes among which DM cannot distinguish.

3 Bayesian Updating and the Bellman Equation

Let $\pi \in [0,1]$ represent the belief of DM that $\omega = 1$. Expected instantaneous rewards are 0 from choosing the empty set; $\lambda^0\pi - c$ from choosing $\{x_0\}$; $\lambda^0(1-\pi) - c$ from choosing $\{x_1\}$; $\lambda^0 - 2c$ from choosing X. Assume that $\lambda^0 > 2c$; even in the absence of learning, the full bundle is more profitable than the empty bundle. Future payoffs are discounted by $\delta \in (0,1)$.

The prior of the DM that $\omega = 1$ is denoted by $\pi_0 \in (0,1)$; her corresponding posterior at the beginning of period t is denoted by π_t. The event of arrival makes the posterior jump to 1, if the arrival is for x_0, or to 0, if the arrival is from x_1. If no arrival results from spending a fraction α_t of time on x_0 over the period $[t, t+\Delta t)$, the posterior becomes:

$$\pi_{t+\Delta t} = \frac{\pi_t e^{-\alpha_t \lambda^0 \Delta t}}{\pi_t e^{-\alpha_t \lambda^0 \Delta t} + 1 - \pi_t}.$$

As Δt shrinks, we obtain $\dot{\pi}_t = -\alpha_t \lambda^0 \pi_t(1-\pi_t)$. If no arrival results from spending a fraction β_t of time on x_1, we get $\dot{\pi}_t = \beta_t \lambda^0 \pi_t (1-\pi_t)$. Finally, by spending time on both tools, either nothing new is learned or the model uncertainty is resolved immediately.

The problem is stationary, and the state is $\pi \in [0,1]$. The state space would still be $[0,1]$ if ω can take interior values, as long as it can take only two possible values $0 < \underline{\omega} < \overline{\omega} < 1$. However, as soon as we move to a larger set of states of nature, the state space of the problem becomes higher-dimensional.

Expected immediate rewards from allocating a unit of time according to $\alpha \in A$ are:

$$\begin{aligned} I(\alpha, \pi, dt) :&= \lambda^0 \left[\alpha_1 \pi + \alpha_2(1-\pi) + \alpha_3\right] dt - (\alpha_1 + \alpha_2 + 2\alpha_3)\, c, \\ &= \alpha_1 \left(\lambda^0 \pi - c\right) dt + \alpha_2 \left(\lambda^0 (1-\pi) - c\right) dt + \alpha_3(\lambda^0 - 2c) dt. \end{aligned}$$

Let $w : [0,1] \to \mathbb{R}$ denote the (optimal, average) value function. The expected continuation value from allocating time according to $\alpha \in A$ is:

$$\begin{aligned} &C(\alpha, \pi, dt) := (\alpha_1 + \alpha_3)\, \lambda^0 \pi \, dt \, w\,(1) + (\alpha_2 + \alpha_3)\, \lambda^0 (1-\pi) dt \, w\,(0) \\ &+ \left(1 - (\alpha_1 + \alpha_3)\lambda^0 - (\alpha_2 + \alpha_3)\lambda^0(1-\pi) dt\right) \left[w(\pi) + (\alpha_2 - \alpha_1)\lambda^0 \pi (1-\pi) w'(\pi) dt\right], \end{aligned}$$

where $w(0) = w(1) = \lambda^0 - c$.

Thus, the Bellman equation of the problem is:

$$w(\pi) = \max_{\alpha \in A} \left\{ \delta I(\alpha, \pi, dt) + e^{-\delta dt} C(\alpha, \pi, dt) \right\}.$$

By invoking the approximations $e^{-\delta dt} \approx 1 - \delta dt$ and $(dt)^n \approx 0$ for all naturals $n \geq 2$ and rearranging terms, we can rewrite the Bellman equation as:

$$\begin{aligned} w(\pi) = \max_{\alpha \in A} \Bigg\{ & \alpha_1 \left[\lambda^0 \pi - c + \frac{\lambda^0 \pi \left[w(1) - w(\pi) \right] - \lambda^0 \pi (1-\pi) w'(\pi)}{\delta} \right] \\ & + \alpha_2 \left[\lambda^0 (1-\pi) - c + \frac{\lambda^0 (1-\pi) \left[w(0) - w(\pi) \right] + \lambda^0 \pi (1-\pi) w'(\pi)}{\delta} \right] \\ & + \alpha_3 \left[\lambda^0 - 2c + \frac{\lambda^0(\lambda^0 - c) - w(\pi)}{\delta} \right] \Bigg\}. \end{aligned}$$

4 Optimal Strategy

Since the expression in braces in the Bellman equation is linear in α, optimal strategies will involve spending the full unit of time on the most promising bundle. Information is more valuable when the DM is sufficiently unsure about the true state. I look for an optimal cutoff strategy $\alpha^* : [0,1] \to A$ with the following properties:

- There is some $\underline{\pi} \in (0,1)$ such that, for all $\pi \in [0, \underline{\pi})$, $\alpha^*(\pi) = (0,1,0)$.
- There is some $\overline{\pi} \in (0,1)$, $\overline{\pi} > \underline{\pi}$, such that, for all $\pi \in (\overline{\pi}, 1]$, $\alpha^*(\pi) = (1,0,0)$.
- For all $\pi \in (\underline{\pi}, \overline{\pi})$, $\alpha^*(\pi) = (0,0,1)$.

Under such strategy α^*, on $(0, \underline{\pi})$, we have:

$$-\lambda^0 \pi (1-\pi) w'(\pi) + (\delta + \lambda^0 (1-\pi)) w(\pi) = \lambda^0 (1-\pi)(\delta + \lambda^0 - c) - \delta c.$$

This equation is similar to Equation (1) in Keller and Rady (2010). The homogeneous part of the solution is $w^H(\pi) := \pi \rho(\pi)^{-\frac{\delta}{\lambda^0}}$, where $\rho(\pi) = \frac{1-\pi}{\pi}$. Notice that

$$w^{H\prime}(\pi) = \frac{1 - \pi + \frac{\delta}{\lambda^0}}{\pi(1-\pi)} w^H(\pi).$$

For the particular part of the solution, we guess and verify an affine function $w^P(\pi) = a(1-\pi) + b$. For this guess to be correct, we must have:

$$\lambda^0 \pi (1-\pi) a + (\delta + \lambda^0 (1-\pi))(a(1-\pi) + b) = \lambda^0 (1-\pi)(\delta + \lambda^0 - c) - \delta c.$$

This gives $b = -c$ and $a = \lambda^0$. Up to a constant of integration C_1, the solution is $w(\pi) = C_1 \pi \rho(\pi)^{-\frac{\delta}{\lambda^0}} + \lambda^0 (1-\pi) - c$.

On $(\overline{\pi}, 1)$, we have:

$$\lambda^0\pi(1-\pi)w'(\pi) + (\delta + \lambda^0\pi)w(\pi) = \lambda^0\pi(\delta + \lambda^0 - c) - \delta c.$$

This equation is almost identical to Equation (1) in Keller and Rady (2010); up to a constant of integration C_0, the solution is

$$w(\pi) = C_0(1-\pi)\rho(\pi)^{\frac{\delta}{\lambda^0}} + \lambda^0\pi - c.$$

Finally, on $(\underline{\pi}, \overline{\pi})$, we have:

$$(\lambda^0 + \delta)w(\pi) = \delta(\lambda^0 - 2c) + \lambda^0(\lambda^0 - c);$$

solve for $w(\pi)$ to get $w(\pi) = \lambda^0 - c - \frac{\delta c}{\lambda^0 + \delta}$.

To identify a specific candidate for an optimal strategy, we need to pin down the thresholds and the constants of integration. We do so by means of the value-matching (VM) and smooth-pasting (SP) conditions.

Condition (VM)**.** $w(\underline{\pi}) = \lambda^0 - c - \frac{\delta c}{\lambda^0 + \delta} = w(\overline{\pi})$

Condition (SP)**.** $w'(\underline{\pi}) = 0 = w'(\overline{\pi})$

The first equality in Condition (VM) is:

$$C_1\underline{\pi}\rho(\underline{\pi})^{-\frac{\delta}{\lambda^0}} + \lambda^0(1-\underline{\pi}) - c = \lambda^0 - c - \frac{\delta c}{\lambda^0 + \delta},$$

which gives:

$$C_1 = C_1(\underline{\pi}) := \frac{\lambda^0(\lambda^0 + \delta)\underline{\pi} - \delta c}{\underline{\pi}(\lambda^0 + \delta)}\rho(\underline{\pi})^{\frac{\delta}{\lambda^0}}.$$

The second equality is very similar to the first, and gives:

$$C_0 = C_0(\overline{\pi}) := \frac{\lambda^0(\lambda^0 + \delta)(1-\overline{\pi}) - \delta c}{(1-\overline{\pi})(\lambda^0 + \delta)}\rho(\overline{\pi})^{-\frac{\delta}{\lambda^0}}.$$

The first equality in Condition (SP) is:

$$\frac{1 - \underline{\pi} + \frac{\delta}{\lambda^0}}{\underline{\pi}(1-\underline{\pi})}\left(\lambda^0\underline{\pi} - \frac{\delta c}{\lambda^0 + \delta}\right) - \lambda^0 = 0;$$

solve for $\underline{\pi}$ to get:

$$\underline{\pi} = \frac{\lambda^0 + \delta}{\lambda^0 + \delta + c} \frac{c}{\lambda^0} \in (0,1).$$

With this expression for $\underline{\pi}$, we can write $C_1(\underline{\pi})$ as $C_1(\underline{\pi}) = \frac{\lambda^0 c}{\lambda^0 + \delta} \rho(\underline{\pi})^{\frac{\delta}{\lambda^0}+1}$. The second equality in Condition (SP) is analogous to the first, and solving for $\overline{\pi}$ gives:[2]

$$\overline{\pi} = \frac{(\lambda^0 + \delta)(\lambda^0 - c) + \lambda^0 c}{\lambda^0(\lambda^0 + \delta + c)} = 1 - \underline{\pi} \in (0,1).$$

This expression allows us to write $C_0(\overline{\pi})$ as $C_0(\overline{\pi}) = \frac{\lambda^0 c}{\lambda^0 + \delta} \rho(\overline{\pi})^{-\frac{\delta}{\lambda^0}-1}$.

Putting all of the pieces together, the solution candidate is:

$$w^0(\pi) = \begin{cases} \frac{\lambda^0 c \rho(\underline{\pi})}{\lambda^0+\delta} \pi \left(\frac{\rho(\pi)}{\rho(\underline{\pi})}\right)^{-\frac{\delta}{\lambda^0}} + \lambda^0(1-\pi) - c & \pi \in [0, \underline{\pi}); \\ \lambda^0 - c - \frac{\delta c}{\lambda^0+\delta} & \pi \in [\underline{\pi}, \overline{\pi}]; \\ \frac{\lambda^0 c}{(\lambda^0+\delta)\rho(\overline{\pi})} (1-\pi) \left(\frac{\rho(\pi)}{\rho(\overline{\pi})}\right)^{\frac{\delta}{\lambda^0}} + \lambda^0 \pi - c & \pi \in (\overline{\pi}, 1]. \end{cases}$$

This function is continuously differentiable, strictly decreasing on $[0, \underline{\pi})$, and strictly increasing on $(\overline{\pi}, 1]$ (See Lemma A1 in the appendix). Figure 2 shows the plot of w for the case $\lambda^0 = 0.7$, $c = 0.3$, and $\delta \in \{0.9, 0.99\}$.

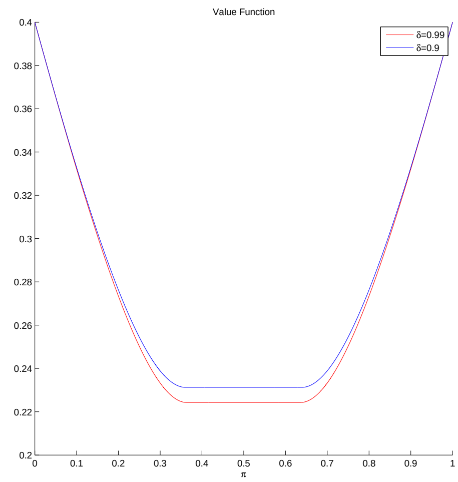


Figure 2: Graph of w^0; $\lambda^0 = 0.7$, $c = 0.3$, and $\delta \in \{0.9, 0.99\}$.

[2] We have $\overline{\pi} > \underline{\pi}$ if and only if $\lambda^0 - 2c > -\frac{\lambda^0 c}{\lambda^0+\delta}$, which is true under the assumption that $\lambda^0 > 2c$.

Since $\overline{\pi} = 1 - \underline{\pi}$, and since these two cutoffs determine the integration constants, we can summarize the cutoff-strategy solution candidate by the lower cutoff, $\underline{\pi}$. Figure 3 plots this cutoff as a function of c, for $\lambda^0 = 0.7$ and $\delta \in \{0.9, 0.99\}$.

The next theorem states that the solution candidate is indeed a solution. The proof is in the appendix.

Theorem 1. *Assume that $\lambda^0 > 2c$. The function w^0 solves the Bellman equation. Thus, the cutoff strategy identified by lower cutoff $\underline{\pi} = \frac{\lambda^0+\delta}{\lambda^0+\delta+c}\frac{c}{\lambda^0} \in (0,1)$ is an optimal strategy.*

5 Conclusion

This paper analyzes a continuous-time variation of the problem described in Francetich and Kreps (2013). A decision maker can choose up to two options over time. Only one of this alternatives is valuable, but the decision makes does not observe which one. I characterize the optimal cutoff strategy.

The structure of the problem studied here is extremely simple. However, while some directions of extensions are feasible (for example, allowing for asymmetries in costs), the problem can become intractable or very cumbersome very quickly. For instance, allowing for a larger set of tools would require expanding the dimensionality of the state space. While some work has been done dealing with vector-valued states (see Klein and Rady, 2011), the limitations to this direction are substantial. Thus, Francetich and Kreps (2013) explores an alternative direction: *heuristics*.

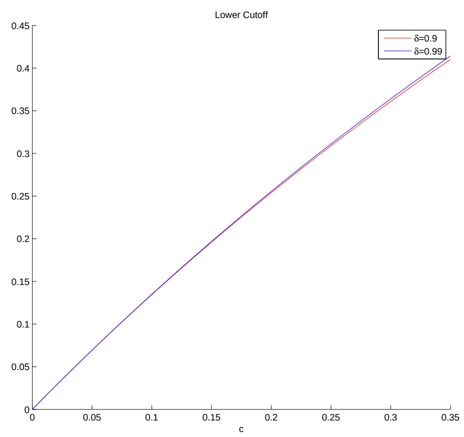


Figure 3: Lower cutoff $\underline{\pi}$ as a function of c, given $\lambda^0 = 0.7$ and $\delta \in \{0.9, 0.99\}$.

References

Bergemann, D. and Valimaki, J. (2001). Stationary multi-choice bandit problems. *Journal of Economic Dynamics and Control* 25:1585–1594.

Francetich, A. and Kreps, D. (2013). Choosing a good toolkit: An essay in behavioral economics. Working Paper.

Keller, G. and Rady, S. (2010). Strategic experimentation with poisson bandits. *Theoretical Economics* 5:275–311.

Klein, N. and Rady, S. (2011). Negatively correlated bandits. *Review of Economics Studies* 78:693–732.

A Proofs

Lemma A1. *The function* $w^0 : [0,1] \to \mathbb{R}$ *given by*

$$w^0(\pi) = \begin{cases} \frac{\lambda^0 c\rho(\underline{\pi})}{\lambda^0+\delta}\pi \left(\frac{\rho(\pi)}{\rho(\underline{\pi})}\right)^{-\frac{\delta}{\lambda^0}} + \lambda^0(1-\pi) - c & \pi \in [0,\underline{\pi}); \\ \lambda^0 - c - \frac{\delta c}{\lambda^0+\delta} & \pi \in [\underline{\pi},\overline{\pi}]; \\ \frac{\lambda^0 c}{(\lambda^0+\delta)\rho(\overline{\pi})}(1-\pi)\left(\frac{\rho(\pi)}{\rho(\overline{\pi})}\right)^{\frac{\delta}{\lambda^0}} + \lambda^0\pi - c & \pi \in (\overline{\pi},1]. \end{cases}$$

is continuously differentiable, strictly decreasing on $[0,\underline{\pi})$*, and strictly increasing on* $(\overline{\pi},1]$*.*

Proof. Continuous differentiability follows from value matching and smooth pasting. On $[0,\underline{\pi})$, we have:

$$\begin{aligned} w^{0\prime}(\pi) &= \frac{\lambda^0 c\rho(\underline{\pi})}{\lambda^0+\delta}\left(\frac{\rho(\pi)}{\rho(\underline{\pi})}\right)^{-\frac{\delta}{\lambda^0}}\left(1+\frac{\delta}{\lambda^0(1-\pi)}\right) - \lambda^0 \\ &< \frac{\lambda^0 c\rho(\underline{\pi})}{\lambda^0+\delta}\left(1+\frac{\delta}{\lambda^0(1-\overline{\pi})}\right) - \lambda^0 = 0. \end{aligned}$$

Finally, on $(\overline{\pi},1]$,

$$\begin{aligned} w^{0\prime}(\pi) &= -\frac{\lambda^0 c}{(\lambda^0+\delta)\rho(\overline{\pi})}\left(\frac{\rho(\pi)}{\rho(\overline{\pi})}\right)^{\frac{\delta}{\lambda^0}}\left(1+\frac{\delta}{\lambda^0\pi}\right) + \lambda^0 \\ &> -\frac{\lambda^0 c}{(\lambda^0+\delta)\rho(\overline{\pi})}\left(1+\frac{\delta}{\lambda^0\overline{\pi}}\right) + \lambda^0 = 0. \end{aligned}$$

This conclues the proof. $\square$

Proof of Theorem 1. We want to verify that w^0 solves the Bellman equation; namely, that:

$$w^0(\pi) = \max\Big\{\lambda^0\pi - c + \frac{\lambda^0\pi\left[\lambda^0 - c - w^0(\pi)\right] - \lambda^0\pi(1-\pi)w^{0\prime}(\pi)}{\delta},$$

$$\lambda^0(1-\pi) - c + \frac{\lambda^0(1-\pi)\left[\lambda^0 - c - w^0(\pi)\right] + \lambda^0\pi(1-\pi)w^{0\prime}(\pi)}{\delta},$$

$$\lambda^0 - 2c + \frac{\lambda^0\left[\lambda^0 - c - w^0(\pi)\right]}{\delta}\Big\}$$

for all $\pi \in [0,1]$. To this end, define:

$$\begin{aligned}
R^0_{w^0}(\pi) &:= \lambda^0\pi - c + \frac{\lambda^0\pi\left[\lambda^0 - c - w^0(\pi)\right] - \lambda^0\pi(1-\pi)w^{0\prime}(\pi)}{\delta};\\
R^1_{w^0}(\pi) &:= \lambda^0(1-\pi) - c + \frac{\lambda^0(1-\pi)\left[\lambda^0 - c - w^0(\pi)\right] + \lambda^0\pi(1-\pi)w^{0\prime}(\pi)}{\delta};\\
R^2_{w^0}(\pi) &:= \lambda^0 - 2c + \frac{\lambda^0\left[\lambda^0 - c - w^0(\pi)\right]}{\delta}.
\end{aligned}$$

We must check the following conditions:

1. On $(0,\underline{\pi})$, $R^1_{w^0}(\pi) - R^0_{w^0}(\pi) > 0$ and $R^1_{w^0}(\pi) - R^2_{w^0}(\pi) > 0$.
2. On $(\underline{\pi},\overline{\pi})$, $R^2_{w^0}(\pi) - R^0_{w^0}(\pi) > 0$ and $R^2_{w^0}(\pi) - R^1_{w^0}(\pi) > 0$.
3. Finally, on $(\overline{\pi},1)$, $R^0_{w^0}(\pi) - R^1_{w^0}(\pi) > 0$ and $R^0_{w^0}(\pi) - R^2_{w^0}(\pi) > 0$.

Start with $\pi \in (0,\underline{\pi})$. In this region, we have: $\lambda^0\pi(1-\pi)w^{0\prime}(\pi) = -[\lambda^0(1-\pi) + \delta](\lambda^0 - c - w^0(\pi)) + \lambda^0\delta\pi$. Thus,

$$\begin{aligned}
R^1_{w^0}(\pi) - R^0_{w^0}(\pi) &= \lambda^0(1-2\pi) + \frac{\lambda^0(1-2\pi)\left[\lambda^0 - c - w^0(\pi)\right] + 2\lambda^0\pi(1-\pi)w^{0\prime}(\pi)}{\delta}\\
&= \lambda^0 - \frac{(\lambda^0 + 2\delta)\left[\lambda^0 - c - w^0(\pi)\right]}{\delta}\\
&> \lambda^0 - \frac{(\lambda^0 + 2\delta)\left[\lambda^0 - c - w^0(\underline{\pi})\right]}{\delta}\\
&= \lambda^0 - \frac{\lambda^0 + 2\delta}{\lambda^0 + \delta}c > 0,
\end{aligned}$$

where the first strict inequality follows from the fact that w^0 is strictly decreasing on

$[0, \underline{\pi})$. Similarly,

$$\begin{aligned} R^1_{w^0}(\pi) - R^2_{w^0}(\pi) &= -(\lambda^0\pi - c) - \frac{\lambda^0\pi\left[\lambda^0 - c - w^0(\pi)\right] - \lambda^0\pi(1-\pi)w^{0\prime}(\pi)}{\delta} \\ &= c_0 - \frac{(\lambda^0+\delta)\left[\lambda^0 - c - w^0(\pi)\right]}{\delta} \\ &> c_0 - \frac{(\lambda^0+\delta)\left[\lambda^0 - c - w^0(\underline{\pi})\right]}{\delta} = 0. \end{aligned}$$

Next, consider $\pi \in (\underline{\pi}, \overline{\pi})$. In this region, $w^{0\prime}(\pi) = 0$. Now,

$$\begin{aligned} R^2_{w^0}(\pi) - R^0_{w^0}(\pi) &= \lambda^0(1-\pi)\frac{\lambda^0+\delta+c}{\lambda^0+\delta} - c \\ &> \lambda^0(1-\overline{\pi})\frac{\lambda^0+\delta+c}{\lambda^0+\delta} - c = 0; \\ R^2_{w^0}(\pi) - R^1_{w^0}(\pi) &= \lambda^0\pi\frac{\lambda^0+\delta+c}{\lambda^0+\delta} - c \\ &> \lambda^0\overline{\pi}\frac{\lambda^0+\delta+c}{\lambda^0+\delta} - c = 0. \end{aligned}$$

Finally, take $\pi \in (\overline{\pi}, 1]$; now, we have $\lambda^0\pi(1-\pi)w^{0\prime}(\pi) = (\lambda^0\pi+\delta)(\lambda^0 - c - w^0(\pi)) - \lambda^0\delta(1-\pi)$. Hence,

$$\begin{aligned} R^0_{w^0}(\pi) - R^1_{w^0}(\pi) &= \lambda^0(2\pi - 1) + \frac{\lambda^0(2\pi-1)\left[\lambda^0 - c - w^0(\pi)\right] - 2\lambda^0\pi(1-\pi)w^{0\prime}(\pi)}{\delta} \\ &= \lambda^0 - \frac{\lambda^0+2\delta}{\delta}\left[\lambda^0 - c - w^0(\pi)\right] \\ &> \lambda^0 - \frac{\lambda^0+2\delta}{\delta}\left[\lambda^0 - c - w^0(\overline{\pi})\right] \\ &= \lambda^0 - \frac{\lambda^0+2\delta}{\lambda^0+\delta}c > 0, \end{aligned}$$

where the strict inequality follows because w^0 is strictly increasing on this region;

$$\begin{aligned} R^0_{w^0}(\pi) - R^2_{w^0}(\pi) &= -[\lambda^0(1-\pi) - c] - \frac{\lambda^0(1-\pi)\left[\lambda^0 - c - w^0(\pi)\right] + \lambda^0\pi(1-\pi)w^{0\prime}(\pi)}{\delta} \\ &= c - \frac{(\lambda^0+\delta)\left[\lambda^0 - c - w^0(\pi)\right]}{\delta} \\ &> c - \frac{(\lambda^0+\delta)\left[\lambda^0 - c - w^0(\overline{\pi})\right]}{\delta} = 0. \end{aligned}$$

This concludes the proof. □