Valuation of the American Put Option: a Dynamic Programming Approach[§]

Lorenzo Garlappi

April 1996

Abstract

The problem of pricing an American put option on a non-dividend-paying stock is analyzed from the perspective of Dynamic Programming. The results obtained are shown to be consistent with the more developed Contingent Claim Analysis paradigm of financial economics literature. A brief investigation of the issues related to continuous time modeling with Ito processes is also considered.

1 Introduction

An option is a security giving the right, but not the obligation, to buy (*call option*) or to sell (*put option*) an asset, for a certain price, within a specified date. The price in the contract is known as *exercise price* or *strike price*; the date in the contract is known as the *expiration date* or *maturity*. A "European option" is one that can be exercised only on a specified future date. An "American option" is one that can be exercised at any time up to the date the option expires.

It is immediate to deduce from the definitions that American-type options involve a more "active" participation of the holder which is called at each time to decide whether or not to exercise his claim, whereas for European options this decision is delayed at the expiration date and assumes the form of a "now-or-never" decision.

Aware of this structure implied by the nature of the American-type options, we will develop in this paper a simple approach for the pricing of American put option based on Dynamic Programming. One may wonder why we explicitly refer to put options and voluntarily exclude call options from our analysis. The answer is a basic well-known result in option pricing according to which it is never optimal to exercise an American call option on a non-dividend paying stock prior to its maturity. In this case, in fact, the value of the American option turns out to coincide with the corresponding European option and well-known techniques for pricing these claims can be used. On the contrary, the above irrelevancy does

[§]Submitted to Martin L. Puterman in partial fulfillment of the requirements for the course COMM 586: "Dynamic Programming and Stochastic Control".

not hold for American put option: it is in fact possible to show that at any time during its life a put option should be exercised if it is sufficiently in the money (i.e. if the price of the underlying share is sufficiently below the exercise price). [1] This implicit "dynamic" content, makes the problem of early exercise of an American put option, even in the simplest case of non-dividend-paying stock, worth investigating.

The decision that the holder of a put option is facing is a timing decision: when is it better to exercise? This problem, whose solution directly leads to the issues of optimal pricing of an option, can be analyzed with two different techniques: *dynamic programming* and *contingent claim analysis*. They are in fact closely related to each other, and lead to identical results in many applications. However they make different assumptions about financial markets and the discount rates that individuals use to value future cash flows.

Dynamic programming is a very general tool for dynamic optimization and is particularly useful in treating uncertainty. It breaks the whole sequence of decisions into just two components: the immediate decision, and a valuation function that encapsulates the consequences of all subsequent optimal decisions, starting with the position that results from the immediate decision. If the planning horizon is finite, as it is in our problem, the very last decision at its end has nothing following it and can therefore be found using standard static optimization methods. This solution then provides the valuation function appropriate to the penultimate decision. That, in turn, serves for the decision two stages from the end, and so on. One can work backwards all the way to the initial condition.

Contingent claim analysis builds on ideas from financial economics. In this approach the put option is seen as an *asset* that has a "dividend" at the exercise date and that, because of this, has some value in the market. The basic assumption is that the pattern of returns generated by the option is always replicable by using a combination of traded assets on the market (this assumption is referred to as *completeness of financial markets*). All one needs is therefore a combination, or portfolio, of traded assets that will replicate the pattern of returns from the option at every future date and in every future uncertain eventuality. The composition of this portfolio needs not be fixed; it could change as the prices of the component assets change. Then the value of the option must be equal to this replicating portfolio, because any discrepancy would present an arbitrage opportunity. In practice, to price an option, a riskless portfolio is formed containing the option and some other "risk-neutralizing" assets; once this portfolio is formed its return is set equal to the return of a riskless bond (risk-free interest rate) and the price of the option is obtained as a consequence.[2]

Although its flexibility and great generality, dynamic programming has not received as much attention as contingent claim analysis from the finance literature, even if the trend of these last years seems moving in the direction of an increasing use of dynamic programming techniques. In this paper we will show how dynamic programming can provide useful results in the problem of pricing an American put on a non-dividend-paying stock. It will be interesting

[1] See Appendix.

[2] This is the logic behind the Black-Scholes (1973) and Merton (1973) model for pricing derivatives securities. Geske and Johnson (1984) provide an analytical formula for the American put derived under the paradigm of contingent claim analysis.

to see that these results can be "dually" read in the perspective of a contingent claim analysis approach. We will also analyze, at an intuitive level, the the problems arising in the case of continuous-time models for pricing the American put and, in the special case of Ito processes, we will provide the structure of the optimal exercise policy.

2 A Two-Period Example

As we mentioned above, an American put option gives the right but not the obligation to sell a particular security at a particular price within a specified date (expiration date). If we call s_t the price of the underlying security on which the option is written and I the exercise price, it turns out that the payoff, V_t, from exercising the put option on one asset at any time t is given by the well-known expression

$$V_t = \max[I - s_t, 0].$$

In order to provide a concrete setting for the problem of valuation of an American put option, let us consider, as a first step, a very simple two-period example. The holder of the put is facing the alternative to either exercise the option now ($t = 0$) or wait one period and decide next period ($t = 1$) whether to exercise or not. To keep things simple let us assume that the price of the underlying security in period $t = 0$ is s_0 and in period $t = 1$ is $s_u \equiv (1 + u)s_0$ with probability q and $s_d \equiv (1 - d)s_0$ with probability $(1 - q)$, where $0 \leq u, d \leq 1$. This is a very simple Markov Decision Problem in which the *decision epochs* are $t = 0$ and $t = 1$; the *states* of the system are represented by the set $\mathcal{S}$ of prices of the underlying security $\mathcal{S} = \{s_0, s_u, s_d\}$; the *actions* at each decision epoch are $\{C, Q\}$ where C means "do not exercise the option" and Q means "exercise the option". The *rewards* at time t are $r_t(s_t, Q) = I - s_t$, if the option is exercised and $r_t(s_t, C) = 0$, if the option is not exercised in t. For simplicity we assume that there are not contracting costs in this last case. Finally, the *transition probabilities* are simply given by the above assumption on the stochastic process for the price of the underlying security: $p(s_u|s_0) = q$ and $p(s_d|s_0) = 1 - q$; we note that these transition probabilities are not affected by the decisions made by the investor and, in this example, are assumed to be time-independent.

The problem faced by the holder of the put option is therefore an optimal stopping problem where stopping means exercise the option. At an intuitive level we can argue that the investor is looking for a strategy (i.e. a stopping time) that maximize the expected present value of his claim. Assuming the existence of an interest rate i at which future payoff are discounted[3], this *optimality criterion* together with the above formulation of the Markov Decision Process, define a finite-horizon *Discounted Markov Decision Problem.* We will see in the next section that this intuitive optimality principle actually has a deep economic meaning and that its validity is supported by exigency of avoiding arbitrage opportunities.

In order to solve the above problem it is natural to work backwards from time 1 to time 0. At time 1 (last decision epoch) the decision to exercise or not will be based on the realization

[3]We interpret this as an opportunity cost for the investor.

of the underlying price at time 1. Let $s_1 \in \{s_d, s_u\}$, if $s_1 < I$ it is convenient to exercise the option, if $s_1 > I$ the option will remain unexercised. Therefore the payoff from an optimal policy at at time 1 is given by:

$$V_1^*(s_1) = \max[I - s_1, 0].$$

At time 0 the investor is facing a different trade-off: exercise now or wait and do what is best when period 1 arrives. To assess this the investor must look ahead to his own actions in different future eventualities. From period 1 onward the condition will not change and so there is no point waiting at time 1 when profitable exercise is possible. Suppose the investor does not exercise at time 0 but instead waits. In period 1 the price of the underlying asset will be s_u with probability q and s_d with probability $1-q$. For each of these two possibilities the investor will exercise if $s_1 > I$, realizing the payoff $V_1^*(s_1)$ determined above. This outcome of future optimal decisions is also called the *continuation value*. From the perspective of period 0, the period-1 price s_1 and therefore the values $V_1^*(s_1)$ are all random variables. Let E_0 be the expectation calculated using the information available at period 0. Then we have:

$$E_0[V_1^*] = q \max[I - s_u, 0] + (1-q) \max[I - s_d, 0]. \tag{1}$$

Now we return to the decision at period 0. The investor has two choices. If he exercises the option, he gets the value $I - s_0$. If he does not, he gets the continuation value $E_0[V_1^*]$ defined above, but that is available in period 1 and hence must be discounted by the factor $1/(1+i)$ to express it in period-0 units. The optimal choice is obviously the one that yields the larger value. Therefore the net present value of our option, optimally managed, is:

$$V_0^* = \max\left\{ I - s_0, \frac{1}{1+i} E_0[V_1^*] \right\} \tag{2}$$

What we unconsciously did in this derivation is simply building up a very simple Bellman equation which gives the solution to the problem of finding the optimal strategy (i.e. optimal stopping time) that maximize the value function V_0^*. In the next section we will prove that this value is the only value for an American put that can "survive" arbitrage strategies in a complete frictionless financial market. It will be interesting to see how a result suggested by the intuition of dynamic programming actually possesses a very deep interpretation in term of equilibrium price in a complete financial market.

The result derived above in an intuitive way captures the essential idea of dynamic programming. We split the whole sequence of decisions into two parts: the immediate choice, and the remaining decision, all of whose effects are summarized in the continuation value. To find the optimal sequence of decisions we work backward. At the last relevant decision point we can make the best choice and thereby find the continuation value (V_1^* in our case). Then at the decision point before that one, we know the expected continuation value and therefore can optimize the current choice. In our example there were just two periods and that was the end of the story. When there are more periods, the same procedure applies repeatedly.

Of course the optimal strategy of the example above varies as long as the parameters s_0, q and i vary. It is particularly interesting to see the sensitivity of the optimal policy to changes in the price of the underlying security. This variable is probably the most important in determining the optimal policy for the holder of a put option: once he observes the price of the underlying security he has to make a decision whether to exercise or not. The intuition is that if the call is sufficiently "in the money", which means that the current price is sufficiently below the exercise price I, than it is optimal to exercise; on the other hand, it seems reasonable to believe that if the option is not sufficiently deep in the money, than it is optimal to wait one period and delay the decision. We will see with a numerical example that this is what actually happen in our case.

EXAMPLE. Let us assume that $u = d = 0.5$, $q = 0.5$, $I = 10$ and $i = 10\%$. We are interested to see how the optimal policy at time 0 changes with the price s_0 of the underlying security at time 0. Substituting the numerical values in (2) and (1) we obtain:

$$\begin{aligned} V_0^* &= \max\left\{10 - s_0, \frac{1}{1.1}E_0[V_1^*]\right\} \\ &= \max\left\{10 - s_0, \frac{1}{1.1}\left[q \max[10 - 1.5s_0, 0] + (1-q)\max[10 - 0.5s_0, 0]\right]\right\} \end{aligned}$$

From which we deduce that

- if $s_0 < 7.058$, then $V_0^* = 10 - s_0$;
- if $7.058 \leq s_0 < 20$, then $V_0^* = \frac{1}{1.1}E_0[V_1^*] = \frac{1}{1.1} \cdot \frac{1}{2}(10 - 0.5s_0)$;
- if $s_0 \geq 20$, then $V_0^* = 0$.

The value of the put option is hence a convex, piecewise-linear function of the current price, s_0 and the optimal decision rule likewise depends on s_0. If $s_0 < 7.058$ the investor should exercise immediately. If $7.058 \leq s_0 < 20$ it is optimal for the investor to wait one period and then exercise in period 1 only if the option is in the money. Finally, if $s_0 \geq 20$, $V_0^* = 0$ and the investor should never exercise. The reason for this last case is to be found in the probabilistic assumptions on the stochastic process governing the price of the underlying security. Given our naive process, if at time 0 the price is greater than 20 it is *almost sure* that the option will end up never being in the money in the next period and therefore the optimal policy suggested is never exercise *a priori*. This solution is summarized in the following table:

Region	**Option Value**	**Optimal Decision Rule**
$s_0 < 7.058$	$10 - s_0$	Exercise
$7.058 \leq s_0 < 20$	$\frac{1}{1.1} \cdot \frac{1}{2}(10 - 0.5s_0)$	Wait
$s_0 > 20$	0	Never Exercise

3 Many Periods

We now generalize the two-period example above to the case of finitely many decision epochs. In this section we continue to assume, however, that uncertainty is modeled using discrete-time Markov processes. In the next sections we will mention how results obtained in this framework can be extended to a continuous time setting where uncertainty takes the form of a Wiener process or more general diffusion processes for the state variables.

Let $\mathcal{S}$ be the set of possible states of the system. At each time t the state of the system is represented by a state variable $s_t \in \mathcal{S}$. In our example this variable was the price of the underlying asset, but, in an very general sense, the theory extends readily to vector states of any dimensions (provided it make sense to write options on a multiple set of state variables). At any date t the current value of this variable s_t is known, but future values $s_{t+1}, s_{t+2} \dots$ are random variables. We suppose that the process is Markov, that is, all the information relevant to the determination of the probability distribution of future values is summarized in the current state s_t.

At each period $t = 1, \dots, N-1$ the investor is facing the problem of exercising the option or wait one period and delay the decision. His action, a_t, at time t, is therefore a binary variable whose value C represent waiting and Q represents exercising at once. We will assume that the value of this variable must be chosen using only the information that is available at that time, namely s_t. In other words we restrict our attention to Markov policies: this is natural in the case of the American put option in which the payoff depend on the underlying price of the asset at the exercise date.[4] It is well known that when, as in our case, rewards and transition probabilities depend on the past only through the the current state of the system, the optimal value functions depend on the history only through the current state of the system and this enable us to insure the existence of optimal policies which depend only on the state of the system at the relevant decision epoch.

The state and the action at any time t affect the immediate reward of the investor and the transition probabilities. We will indicate with $\Phi_t(s_{t+1}|s_t, a_t)$ the cumulative probability distribution function of the state next period conditional upon the current information. If the set of states is discrete (as in the example of the previous section) then $\Phi_t(\cdot)$ represents a transition probabilities. Let τ represent a Markov policy, namely a stopping time, and let $\mathcal{T}$ be the set of possible Markov policies. Every policy τ generates a future stream of payoff whose expected present value is captured by the value function V_0^τ. The problem of finding the value of the American put can hence be reduced to the problem of finding the optimal stopping strategy that maximize the present value of the payoff to the investor. In other words we need to solve the following problem:

$$V_0^* = \max_{\tau \in \mathcal{T}} V_0^\tau \tag{3}$$

Since the rewards in our problem are bounded and the action space is compact (binary decision), a key result in dynamic programming[5] allows us to state that the solution of the

[4] See Puterman (1994), Ch. 4.

[5] See Puterman (1994), Ch. 4.4.

above problem is given by solving recursively the optimality equations:

$$V_t^*(s_t) = \max\left\{I - s_t, \frac{1}{1+i}E_t[V_{t+1}^*(s_{t+1})]\right\} \quad t = 0, 1, \dots N-1 \quad s_t \in \mathcal{S} \tag{4}$$

with boundary condition

$$V_N^*(s_N) = \max\{I - s_N, 0\}, \quad s_N \in \mathcal{S} \tag{5}$$

Equation (4) is the *Bellman Equation* for our problem. A glance at (4) tells us that the result obtained in the previous section, somewhat intuitively, perfectly fits into the theoretical framework of the Bellman's Principle of Optimality:

> "an optimal policy has the property that, whatever the initial action, the remaining choices constitute an optimal policy with respect to the subproblem starting at the state that results from the initial actions." (Bellman, 1957, p.83).

Here the optimality of the remaining choices $a_{t+1}, a_{t+2} \dots a_{N-1}$ is subsumed in the continuation value, so only the immediate action a_t remains to be chosen.

Comparison with Contingent Claim Analysis

In the traditional contingent claim approach in finance, the solution (3) to the Bellman equation is referred to as the *no-arbitrage* price for an American put. In order to see it is assumed that market are complete, that is, the payoff of any new security introduced in the market can be replicated or "synthesized" by using combination of existing securities. In this perspective we can see the American put as a security that pays a "dividend" δ^τ when the investor decides to exercise it (time τ) and nothing when he decides to keep it alive.[6] A solution τ^* to (3) is a *rational exercise policy* for the American put in the sense that it maximizes the initial arbitrage-free value of the security.

The main insight from contingent claim analysis is that the initial arbitrage-free value of V_0 *must* be V_0^*, the value of the security with an optimal exercise strategy. To prove this, suppose, by contradiction, that $V_0^* > V_0$. In this case an investor can buy an American put at price V_0, adopt for it a rational exercise policy (i.e. exercise the put at time τ^*) and replicate the dividend process $-\delta^{\tau^*}$. This is possible because we assumed complete markets and the cost of this replicating strategy is, of course, $-V_0^*$. As a result the investor is making an initial profit of $V_0^* - V_0$ and is perfectly matching the future dividend stream, an arbitrage. Conversely, suppose that $V_0 > V_0^*$. Then one could sell the American put for V_0. By assumption, the dividend process δ^τ generated by this option can be synthesized by a trading strategy whose initial cost is V^τ. Moreover, by (3) and by assumption, $V^\tau < V_0^* < V_0$. Therefore we end up with an initial profit of $V_0 - V_0^\tau > 0$ and no further dividends, an arbitrage.

[6] This view allow a large level of generality in analyzing securities in a financial market: every security can in fact be identified as with the stochastic process of the dividends it generates through time. See Duffie (1995).

Structure of the optimal policy

By looking at a simple property of V_t^* we can provide the structure of the optimal policy in a multi-period setting. The following results represent a straightforward generalization of the structure of the optimal policy shown in section 2.

From equations (4) and (5) it is clear that for some values of s_t termination will be optimal and for other values continuation will be optimal. In our earlier example there was a single cutoff level $s^* = 7.058$ with exercise optimal for $s < s^*$ and non exercise optimal for $s > s^*$. Although it is true that $I - s$ is decreasing in s_t and therefore continuation seems plausible for large values of s_t, for arbitrary specification of $\Phi_t(\cdot|\cdot)$ the shape of the continuation and stopping regions could be any sequence of alternating intervals. However, if the distribution $\Phi_t(\cdot|\cdot)$ exhibit *first-order stochastic dominance*, in the sense that the cumulative probability distribution $\Phi_t(s_{t+1}|s_t)$ of future prices s_{t+1} shifts to the right when the current value of s_t increase, then the optimal policy preserve the "monotonicity" structure we observed in our earlier example. Before proving this we need the following preparatory lemma.

Lemma 3.1. If the cumulative probability distribution function $\Phi_t(\cdot|\cdot)$ exhibits first-order stochastic dominance and the conditional expected value of the increment $E_t[s_{t+1} - s_t]$ is nonincreasing in s_t, then $V_t^*(s_t) + s_t$ is increasing in s_t.

Proof: Since $V_N^*(S_N) = \max\{I - s_N, 0\}$ the result is obvious for $t = N$. Suppose now that $V_t^*(s_t) + s_t$ is increasing for $t+1, t+2, \ldots, N$. Then, from equation (4)

$$V_t^*(s_t) + s_t = \max\left\{I, \frac{1}{1+i}E_t[V_{t+1}^*(s_{t+1}) + s_{t+1}] - \frac{1}{1+i}E_t[s_{t+1} - s_t] + \frac{i}{1+i}s_t\right\}.$$

By the induction hypothesis,

$$V_{t+1}^*(s_{t+1}) + s_{t+1}$$

is increasing in s_{t+1}. Since $\Phi(\cdot|\cdot)$ exhibits first-order stochastic dominance, $E_t[V_{t+1}^*(s_{t+1}) + s_{t+1}] = \int (V_{t+1}^*(s_{t+1}) + s_{t+1}) d\Phi(s_{t+1}|s_t)$ is increasing in s_t. By assumption, $E_t[s_{t+1} - s_t]$ is nonincreasing in s_t and therefore the result follows. □

Remark. We note that, if markets are efficient, i.e. if the discounted prices of underlying assets are a martingale, then the restriction we imposed on the conditional expected value $E_t[s_{t+1} - s_t]$ is satisfied.

The following proposition gives the explicit structure of the optimal policy.

Proposition 3.2. The optimal exercise policy for the American put has the following form: There are increasing values of the underlying prices

$$s_1^* \leq s_2^* \leq \cdots \leq s_N^*$$

such that at time t it is optimal to exercise the option if and only if $s_t \leq s_t^*$

Proof: If the price of the underlying asset at time t is s_t then, from equation (4) it is optimal to exercise the option if

$$V_t^*(s_t) \leq I - s_t.$$

Let

$$s_t^* = \min\{s_t : V_t^*(s_t) = I - s_t\}.$$

By Lemma 3.1 it can be seen that, for $s_t \leq s_t^*$,

$$V_t^*(s_t) + s_t \leq V_t^*(s_t^*) + s_t^* = I$$

which shows that it is optimal to exercise the option if at time t the price s_t of the underlying asset is below the threshold s_t^*.

That s_t^* is increasing in t follows from the fact that V_t^* is nonincreasing in t. This is immediate because having less time to decide whether to exercise or not cannot increase one's expected profit. □

The result of the above theorem can be interpreted as follows: the further away an investor is from the expiration of the option the more "exigent" he is towards the underperformance of underlying securities (i.e. s_1^* is the lowest threshold value). As time of expiration approaches he becomes "less exigent", committing himself to an exercise threshold increasing through time.

Remark. We point out that the condition of first-order stochastic dominance driving the above results is true for most of the stochastic processes assumed in the finance literature: random walks, Brownian motion, mean-reverting autoregressive processes and, indeed, in almost all economic applications we can think of.

4 Continuous Time

We will try to provide in this section an intuitive derivation of the continuous-time version of the above model for pricing an American put. Far from being complete, the analysis aims to focus on the intuition driving the main results.

Let us consider the general finite horizon problem stated above, but suppose each time period is of length Δt. Ultimately we are interested in the limit where Δt goes to zero and time is continuous. Let ρ be the discount rate per unit of time, so that the total discounting over an interval of length Δt is given by the factor $1/(1+\rho\Delta t)$. We rewrite the Bellman equation (4), by explicitly considering V^* a function of the state s and of time t:

$$V^*(s,t) = \max\left\{I - s, \frac{1}{1+\rho\Delta t}E[V(s', t+\Delta t)|s]\right\},$$

where s and s' represent the state of the system at time t and $t+\Delta t$ respectively. As Δt goes to zero one may be tempted to guess that the Bellman equation for our optimal stopping problem becomes:

$$V^*(s,t) = \max\left\{I - s, \frac{1}{1+\rho dt}E[V^*(s+ds, t+dt)|s]\right\}. \tag{6}$$

This passage from discrete to continuous time is very casual and heuristic and it is fair to warn the reader that some quite tricky issues are hidden, and must be handled carefully in more rigorous treatments. In discrete time we stipulated that the action a_t taken in the current period t could depend on the knowledge of the current state s_t, but not on the random future state s_{t+1}. In continuous time the two coalesce. We have to be careful not to allow choices to depend on information about the future, otherwise we would be acting with the benefit of hindsight and could make infinite profits. Technically this is avoided in continuous time model by assuming that the uncertainty is "continuous from the right" in time, while the strategies are "continuous from the left". Then any jumps in the stochastic process occur *at an instant*, while the actions cannot change until *just after the instant.*[7] Moreover, although the passage from Δt to dt seems harmless, a big step has been omitted. The limit on the right-hand side of the above expression in Δt depends on the expectation corresponding to the random state s' at time $t + \Delta t$. The big problem now is the structure of the continuous time stochastic processes that allows such limits in a form conducive to further analysis. A class of process that are particularly useful for many finance applications is represented by *Ito processes*. Ito processes play a crucial role in the finance literature of continuous time; most of the original results in continuous time asset pricing, in fact, have been derived assuming this kind of stochastic processes. Although our aim here is not to investigate properties of these processes, we will try to analyze in the next subsection the structure of our problem when the state variable is assumed to evolve continuously in time according to an Ito process.

Ito Processes

The building block of an Ito process is represented by *Brownian Motion*. Brownian motion —also called *Wiener process*— is a continuous-time stochastic process with three important characteristics. First, it is a *Markov process*, i.e. the probability distribution for all future values of the process depends only on its current value and is unaffected by past values of the process or by other current information. Second, the Brownian Motion has *independent increments*: this means that the probability distribution of the process over any time interval is independent of any other non overlapping time interval. Third, changes in the process over any finite interval of time are *normally distributed* with a variance that increases linearly with the time interval.

These three conditions may seem at first glance quite restrictive for modeling real-world variables such as stock prices. In fact while it probably seems reasonable that stock prices satisfy the Markov property and have independent increments, it is not reasonable to assume that price changes are normally distributed since this can lead to the paradox of having negative stock prices. However, through the use of suitable transformations, the Brownian motion can be used as a building block to model an extremely broad range of variables that vary continuously and stochastically through time.

We now restate the above property of Brownian motion more formally. If $z(t)$ is a Brownian motion, than any change in z, Δz, corresponding to a time interval Δt, satisfies the

[7] See Duffie (1988).

following conditions:

1. The relationship between Δz and Δt is given by

$$\Delta z = \epsilon_t \sqrt{\Delta} t,$$

where ϵ_t is a normally distributed random variable with mean of zero and standard deviation of 1.

2. The random variable ϵ_t is serially uncorrelated, that is, $E[\epsilon_t \epsilon_s] = 0$ for $t \neq s$. Thus the values of Δz for any two different intervals of time are independent. This means that $z(t)$ follows a Markov process with independent increments.

By letting Δ become infinitesimally small, we can represent the increment of Brownian motion, dz, in continuous time as

$$dz = \epsilon_t \sqrt{d} t. \tag{7}$$

Since ϵ_t has zero mean and unit standard deviation, $E[dz] = 0$ and $E[(dz)^2] = dt$. Note, however, that Brownian motion has no time derivative in a conventional sense; $\Delta z / \Delta t = \epsilon_t (\Delta t)^{-1/2}$, which becomes infinite as Δt approaches zero. This last property is at the origin of the theory of "differentiation" for this kind of stochastic process which is known as *Ito calculus* and that plays a crucial role in large part of finance literature of continuous time.

We mentioned earlier that the Brownian motion is the main building block in the construction of the more general *Ito processes*. We say that a random variable s follows an Ito process if its evolution through time is given by the following generalization of expression (7):

$$ds = a(s,t)dt + b(s,t)dz \tag{8}$$

where dz is the increment in the Brownian motion, and $a(s,t)$ and $b(s,t)$ are known (non random) functions of s and t.

Let us consider the mean and variance of the increments of this process. Since $E[ds] = 0$, $E[ds] = a(s,t)$. The variance of ds is equal to $E[(ds)^2] - (E[ds])^2$ which contains terms in dt, in $(dt)^2$ and in $(dz)(dt)$, which is of order $(dt)^{3/2}$. For dt infinitesimally small, terms in $(dt)^2$ and in $(dt)^{3/2}$ can be ignored, and, to order dt, the variance is

$$Var[ds] = b^2(s,t).$$

The term $a(s,t)$ is usually referred to as *drift rate* of the Ito process, while the term $b^2(s,t)$ is the *instantaneous variance rate.*

Referring back to our problem of pricing an American put we will assume that the price of the underlying asset, s, follows an Ito process of the form given in equation (8). A closer look at the Bellman equation (6) for the optimal stopping problem we derived above, tells us that, in order to deal with the case of continuous time, we must be able to take differentials of the value function $V(s,t)$ with respect to s and t. While the differentiation with respect to t does not create problems, the differentiation with respect to s cannot be performed with usual calculus technique. The reason, mentioned above, is that s is a random variable which follows a stochastic process which is *nowhere* differentiable. Fortunately the theory of differentiation for functionals of Ito process is well established and allows us to complete our analysis of the optimal stopping problem. The main result in the theory of differentiation of functionals of Ito process is *Ito's Lemma.*

Ito's Lemma is easiest to understand as Taylor series expansion. Suppose $s(t)$ follows the process of equation (8), and consider a function $V(s,t)$ that is at least twice differentiable in s and once in t. Ito's Lemma gives the differential dV as

$$dV = \frac{\partial V}{\partial s}ds + \frac{\partial V}{\partial t}dt + \frac{1}{2}\frac{\partial^2 V}{\partial s^2}(ds)^2,$$

or, substituting equation (8) for ds:

$$dV = \left[\frac{\partial V}{\partial t} + a(s,t)\frac{\partial V}{\partial s}ds + \frac{1}{2}b^2(s,t)\frac{\partial^2 V}{\partial s^2}\right]dt + b(s,t)\frac{\partial V}{\partial s}dz \tag{9}$$

We are now ready to provide the continuous version of the optimal stopping strategy derived in section 3. In analogy with the above results we can state that, in a continuous framework, the optimal policy will be characterized by a *time-varying critical threshold* $s^*(t)$ for the stock price. We can interpret the critical values $s^*(t)$ for various t as forming a curve that divides the (s,t) space into two regions, with continuation (no exercise of the option) optimal above the curve and stopping (exercise of the option) optimal below it. Of course we do non know the equation of the curve $s = s^*(t)$ in advance, but must find it out as part of the solution of the dynamic programming problem. This problem is referred to in the predominant literature as a *free-boundary problem.*

The Bellman equation for our optimal stopping problem is (6), which we repeat here for ease of reference:

$$V^*(s,t) = \max\left\{I - s, \frac{1}{1+\rho dt}E[V^*(s+ds, t+dt)|s]\right\}.$$

In the continuation region, the second term on the the right-hand side is the the larger of the two. Applying Ito's Lemma to the value function V^* we obtain:

$$E[V^*(s+ds,t+dt)|s] = V^*(s,t) + \left[\frac{\partial V^*}{\partial t} + a(s,t)\frac{\partial V^*}{\partial s}ds + \frac{1}{2}b^2(s,t)\frac{\partial^2 V^*}{\partial s^2}\right]dt + o(dt)$$

where we used the fact that $E[dz] = 0$. Using the above expression in the Bellman equation we can easily derive the following partial differential equation (valid in the continuation region) that the value function must satisfy:

$$\frac{1}{2}b^2(s,t)\frac{\partial^2 V^*}{\partial s^2} + a(s,t)\frac{\partial V^*}{\partial s} + \frac{\partial V^*}{\partial t} - \rho V^* = 0. \tag{10}$$

As mentioned, this holds for $s > s^*(t)$, and we must look for boundary conditions that hold along $s = s^*(t)$. From the Bellman equation, we know that in the stopping region we have $V^*(s,t) = I - s$, therefore, by continuity, we can impose the condition

$$V^*(s^*(t),t) = I - s^*(t) \quad \forall t \tag{11}$$

This is often called the *value-matching condition* since it matches the values of the unknown function $V^*(s,t)$ to those of the known termination payoff function $I - s$. An intuitive explanation of (11) proceeds as follows. Suppose, by contradiction that $V^*(s^*(t), t) < I - s^*(t)$. By continuity, we will have $V^*(s,t) < I - s$ for s just slightly to the right of $s^*(t)$. By Ito's Lemma, for sufficiently small dt the continuation value in the Bellman equation (6) will be less than $I - s$ for s slightly greater than $s^*(t)$. Then, immediate stopping will be optimal for such s, contrary to the definition of $s^*(t)$. The argument for $V^*(s^*(t), t) < I - s^*(t)$ proceeds similarly.

As we mentioned, the boundary $s^*(t)$ is itself unknown, in other words, the region in the (s,t)-space over which the partial differential equation (10) is valid is itself endogenous. It is therefore clear that we need a second condition in addition to (11) if we are to find $s^*(t)$ jointly with the function $V^*(s,t)$. The general mathematical theory of partial differential equation is of little help here and the conditions applicable to free boundaries are specific for each application. In our case the right condition requires that for each t, the values $V^*(s,t)$ and $I - s$, regarded as function of s should meet tangentially at the boundary $s^*(t)$, or

$$\left.\frac{\partial V^*(s,t)}{\partial s}\right|_{s=s^*(t)} = 1 \tag{12}$$

This is called the *smooth-pasting condition* because it requires not just the values but also the derivatives of slopes of the two functions to match at the boundary. An intuitive explanation of this condition can be obtained by thinking of the payoff schedule of a put option as a function of the underlying price s. If, by contradiction, the functions $I - s$ and $V^*(s,t)$ do not meet tangentially in $s^*(t)$ then they must form a "downward-pointing" kink.[8] If this were the case, $s^*(t)$ cannot be a point of indifference between the choice of exercise the option or waiting. It can be proved, in fact, that waiting for a short interval of time Δt is definitely the better policy. The intuitive idea is that by waiting a little bit longer, the investor can observe the next step of the price s and choose position on each side of the kink. This means that the value now of the position that the investor will take in Δt instants is the expected value of the positions on either sides of the kink. Since the expected value is a linear operator, the average of the two positions does better than the kink point itself. This is true even though this average must be discounted because it occurs Δt instant later. The reason is that, as we mentioned, for Brownian motion the steps Δs are proportional to $\sqrt{\Delta} t$, and so is their effect on the value, while the effect of discounting is proportional to Δt. When Δt is small the former effect is relatively much larger and we obtain the required contradiction.[9]

Equation (10), together with the value-matching and smooth-pasting conditions, completes the description of the problem of the value of the American put option in a continuous time framework when the underlying asset price follows an Ito process. The time-varying threshold $s^*(t)$ turns out to be an increasing function of t, in perfect analogy with the result derived in proposition 3.2.[10] It is important to highlight that, although the above derivation

[8] It may be helpful to think of the example in section 2.

[9] For a formal derivation of these arguments, see Dixit (1991).

[10] See Duffie (1995).

required a tedious analysis of formal details, the intuition and the logic of the solution is the same we followed in the simple case of a discrete time two-period model of section 2. This section should also warn the reader of the level of complexity to which this problem can lead when we assume alternative stochastic processes for which there does not exist a richness of results as in the case of Ito processes. The analysis of these issues is however outside the purpose of this paper.

5 Conclusions

In this paper we presented a Dynamic Programming approach to the pricing of an American put option. Starting from a very simple example we have shown how the basic concepts of Dynamic Programming can be extensively interpreted under the perspective of the more traditional Contingent Claim Analysis approach to pricing derivative securities. In particular it has been shown how the solution of the Bellman equation defining the optimal stopping problem for an American put can be directly interpreted as the arbitrage-free value of this claim that would prevail in a complete financial market.

The Dynamic programming approach to pricing derivatives securities is not new in the finance literature but is definitely less developed that the more exploited Contingent Claim Analysis approach. The extreme flexibility and the naturalness with which the main results in this paper have been derived leaves room for an extensive application of these technique in finance. Indeed, the trend in ongoing research seems to be in this direction.

We believe that the theory of Markov decision process can provide interesting insights in the solution of many existing puzzle in theoretical finance. Setting aside the natural extensions to the above results to the case of dividend-paying securities and transaction costs, we believe that one of the most interesting directions for further research is the whole literature concerning "exotic" options. These class of options are characterized by the fact that the payoff at the exercise date may depend on some functionals of the history of the underlying price. By a suitable definition of the state space we think it is possible to embed the problem of pricing exotic options in the class of Markov decision problems which can capture, with sufficient flexibility, the essence of the pricing problem for these instruments. There is apparently no evidence in the literature of any attempt to deal with these problems under the perspective of Markov Decision Processes and we believe that, although it is challenging and with highly uncertain outcome, this is definitely an exciting research topic worth investigating.

Nothing original has been derived in this paper, however we hope that the exercise of "looking things from a different perspective" has been useful to unveil one of the many fields in which Dynamic Programming can provide a substantial and persistent contribution.

Appendix

We proof that it is never optimal to exercise early an American call option on a non-dividend-paying stock. Consider the following two portfolios at time t:

- *Portfolio A*: one American call option on one share with exercise price I plus a discount bond that will be worth I at time T
- *Portfolio B*: one share with value S.

Let $i > 0$ be the period interest rate. The value of the discount bond at time t is $I \cdot (1+i)^{-(T-t)}$. If the call option were exercised at time t, the value of portfolio A would be

$$S - I + I \cdot (1+i)^{-(T-t)}.$$

This is always less than S when $t < T$ since $i > 0$. Portfolio A is therefore worth less than portfolio B if the call option is exercised prior to maturity. If the call option is held to maturity, the value of portfolio A at time T is

$$\max[0, S_T - I] + I = \max[S_T, I]$$

where S_T is the price of the stock a time T. The value of portfolio B at time T is S_T. As long as there is some chance that $S_T < I$, portfolio A is always worth at least as much as portfolio B.
Therefore, portfolio A is worth less than portfolio B if the option is exercised immediately, but it is worth at least as much as portfolio B if the holder of the option delays exercise until maturity. It follows that a call option on a non-dividend-paying stock should never be exercised prior to maturity and its value corresponds to the value of an European call option on the same stock.

The argument above does not old for American put options on non-dividend-paying stocks. Indeed, at any time during its life a put option should be exercised if it is sufficiently in the money. Consider the following two portfolios:

- *Portfolio C*: one American put option on one share with exercise price I plus one share with value S
- *Portfolio D*: a discount bond that will be worth I at time T

If the option is exercised at time $t < T$ portfolio C becomes worth I while portfolio D is worth $I \cdot (1+i)^{-(T-t)}$. Portfolio C is therefore worth more than portfolio D. If the put option is held to maturity, portfolio C becomes worth

$$\max[I - S_T, 0] + S_T = \max[I, S_T]$$

while portfolio D is worth I. Portfolio C is therefore worth at least as much as, ad possibly more (if the put option is exercised early), than portfolio D.

References

[1] Bellman, R. E. (1957). *Dynamic Programming.* Princeton University Press, Princeton, NJ.

[2] Dixit, Avinash. (1991). "A Simplified Treatment of the Theory of Optimal Regulation of Brownian Motion". *Journal of Economic Dynamics and Control* 15, 657-673.

[3] Duffie, Darrell. (1988). *Securities Markets: Stochastic Models.* San Diego, CA: Academic Press.

[4] Duffie, Darrell. (1995). *Dynamic Asset Pricing Theory.* Princeton, NJ: Princeton University Press.

[5] Geske, Robert and H.E. Johnson. (1984). "The American Put Option Valued Analytically". *Journal of Finance* 5, 1511-1524.

[6] Hull, John (1989). *Options, Futures, and Other Derivatives Securities.* Prentice-Hall International, Inc.

[7] Merton, Robert (1973). "The Theory of Rational Option Pricing". *Bell Journal of Economics and Management Science* 4, 141-183.

[8] Puterman, Martin L. (1994). *Markov Decision Processes.* John Wiley & Sons, Inc.

[9] Ross, Sheldon M. (1983). *Introduction to Stochastic Dynamic Programming.* Academic Press, New York.