1
Introduction

In situations where action entails a fixed cost, optimal policies involve doing nothing most of the time and exercising control only occasionally. Interest in economic models that exhibit this type of behavior has exploded in recent years, spurred by growing evidence that "lumpy" adjustment is typical in a number of important economic settings.

For example, the short-run effects of monetary policy are connected with the degree of price stickiness. Data from the U.S. Bureau of Labor Statistics on price changes at retail establishments for the period 1988–2003 suggest that price adjustment is sluggish, at least for some products. The average duration of a price varies greatly across different types of products, with some—gasoline, airfares, and produce—displaying frequent price changes, and others—taxi fares and many types of personal services—displaying much longer durations. The average duration across all products is about 5 months, but the range is wide, from 1.5 months at the 10th percentile to 15 months at the 90th percentile. The size distribution of adjustments is perhaps even more informative about price stickiness. Many adjustments are large, even after short-term sale prices have been removed. The average size is more than 8%. Moreover, 1988–2003 was a period of low inflation, so in the broad sample 45% of price changes are negative, and price cuts are almost as large as price increases. More detailed data for the largest three urban areas (New York, Los Angeles, and Chicago), show that about 30% of price changes are greater than 10% in magnitude, and the figure is about the same for increases and decreases.

The fact that many price changes are large suggests that there may be substantial fixed costs associated with changing a price. Otherwise it is difficult to explain why the changes are not carried out in series of smaller, more frequent increments.

Investment behavior provides a second example. Establishment-level data from the U.S. Census Bureau on 13,700 manufacturing plants for the period 1972–1988 show lumpy adjustment in two ways. First, more than half of the plants in the sample display an episode of large adjustment, at least 37%. In addition, a substantial fraction of aggregate investment, 25%, is concentrated in plants that are increasing their capital stock by more than 30%. At the other end of the distribution, over half of plants increase their capital stock by less than 2.5%, accounting for about 20% of total investment. Thus, for individual plants, changes in size come in substantial part from large one-year increases. From the aggregate point of view, a sizable share of total investment is devoted to these episodes of large increase. As with the price data, the evidence on investment suggests that fixed costs are important.

Evidence on job creation and destruction displays patterns similar to those in the investment data. The opening of a new establishment, the expansion of an old one, or the addition of a shift leads to concentrated job creation, while plant closings and mergers lead to concentrated destruction. Data from the U.S. Census Bureau on 300,000–400,000 manufacturing plants for the period 1972–88 show that two-thirds of total job creation and destruction occurs at plants where employment expands or contracts by more than 25% within a twelve-month period. One-quarter of job destruction arises from plant closings. Again, these adjustment patterns suggest that fixed costs are important.

For consumers, fixed costs are clearly important for durable goods such as housing and automobiles. U.S. Census Bureau data for 1996 show that among individuals aged 15 and older living in owner-occupied housing, the median tenure at the current residence is eight years, and about 40% have been at their current residence for more than eleven years. Although individuals have many motives for staying in their current residences, the substantial transaction costs—including time costs—involved in buying, selling, and moving clearly make frequent moves unattractive.

Evidence suggests that fixed costs are also important for explaining individual portfolio behavior. Absent fixed costs, it is hard to understand why so many households fail to participate in the stock market. The fact that lagged stock market participation increases the likelihood of current participation very strongly—by 32 percentage points, even after controlling for age, education, race, income, and other factors—strongly suggests that fixed entry costs are important. In addition, the fact that wealthier households are more likely both to own stocks and to trade suggests that fixed costs per period or per transaction are also important. In this context and others the costs of gathering and processing information are likely to be important components of the fixed cost.

These examples suggest that models using a representative firm and representative household may be inadequate for studying some questions. For example, the economic effects of an aggregate productivity shock or aggregate demand shock depend on the investment and hiring/firing response of firms. If fixed costs are important, describing the aggregate response requires averaging over a group of firms that are doing little or nothing and a group making substantial adjustments. Although it may happen that movements in the aggregate can be generated by a representative agent, it is difficult to confirm this—or to determine what the representative agent should look like—without explicitly carrying out the aggregation. Moreover, a representative agent that serves as an adequate proxy during periods of calm may be misleading during episodes when the economic environment becomes more turbulent. And on the household side, explicitly taking heterogeneity into account may have a substantial impact on conclusions about welfare.

In situations where fixed costs are important, continuous-time models in which the stochastic shocks follow a Brownian motion or some other diffusion have strong theoretical appeal. An optimal policy in this type of setting involves taking action when the state variable reaches or exceeds appropriately chosen upper and/or lower thresholds and doing nothing when the state lies inside the region defined by those thresholds. Continuous-time models permit a very sharp characterization of the thresholds that trigger an adjustment and the level(s) to which adjustment is made. Indeed, the thresholds and return points can often be characterized as the solution to a system of three or four equations in as many unknowns. The goal in this book is to develop the mathematical apparatus for analyzing models of this type. In the rest of this introduction the structure of typical models is briefly described and a few examples are discussed.

Suppose that in the absence of control the increments to a state variable $X(t)$ are those of a Brownian motion. The (flow) return to the decision maker at any date, $g(X(t))$, depends on the current state, where the function g is continuous and single peaked. Suppose the peak is at $x = a$, so that g is increasing on $(-\infty, a)$ and decreasing on $(a, +\infty)$. The decision maker can adjust the state by discrete amounts, and there is a fixed cost c associated with making any adjustment. For now suppose that the fixed cost is the only cost of adjustment. The decision maker's objective is to maximize the expected discounted value of returns net of adjustment costs, where future returns and costs are discounted at a constant interest rate r. The standard menu cost model has this structure, with $X(t)$ interpreted as the firm's price (in log form) relative to an industrywide or economywide price index that fluctuates stochastically.

The problem for the decision maker is to balance two conflicting goals: maintaining a high return flow by keeping the state in the region around a and avoiding frequent payment of adjustment costs. An optimal policy in this setting has the following form: the decision maker chooses threshold values b, B for the points where control will be exercised, with $b < a < B$, and a return point $S \in (b, B)$ to which the state is adjusted when control is exercised. As long as the state remains inside the open interval (b, B)—called the *inaction region*—the decision maker exercises no control: adjustment is not required. When the state falls to b or rises to B the fixed cost is paid and the state variable is adjusted to the return point S. If the initial state lies outside (b, B) an adjustment to S is made immediately. If the fixed cost c is sufficiently large relative to the range of the return function g, it is possible that $b = -\infty$, $B = +\infty$, or both.

Notice that in general an optimal policy does not involve returning to the point a where instantaneous returns are at a maximum. That is, in general $S \neq a$. For example, if the drift is positive, $\mu > 0$, the decision maker might choose $S < a$, anticipating that (on average) the state will rise. Or, if the return function g is asymmetric around a, he might skew the return point in the direction of higher returns.

Let $v(x)$ denote the expected discounted return from following an optimal policy, given the initial state $X(0) = x$. The first step in finding the optimum is to formulate a Bellman equation involving the function v. That equation can then be used to characterize the optimal policy. Indeed, it can be used in two different ways.

Suppose that thresholds b and B have been selected and that the initial state $X(0) = x$ lies between them, $b < x < B$. Let $\mathrm{E}_x[\cdot]$ and $\mathrm{Pr}_x[\cdot]$ denote expectations and probabilities conditional on the initial state x, and define the random variable $T = T(b) \wedge T(B)$ as the first time the process $X(t)$ reaches b or B. T is an example of a *stopping time*. It is useful to think of $v(x)$ as the sum of three terms:

$$\begin{aligned} v(x) = {} & \text{expected returns over } [0, T) \\ & + \text{expected returns over } [T, +\infty) \text{ if } b \text{ is reached before } B \\ & + \text{expected returns over } [T, +\infty) \text{ if } B \text{ is reached before } b. \end{aligned}$$

Let $w(x, b, B)$ denote the first of these terms, the expected returns up to the stopping time T. The key to making this problem tractable is the fact that $w(x, b, B)$, which is the expected value of an integral over time up to the stopping time $T = T(b) \wedge T(B)$, can be written as an integral over states in the interval $[b, B]$. Specifically,

$$\begin{aligned} w(x, b, B) &\equiv \mathrm{E}_x \left[\int_0^T e^{-rt} g(X(t)) dt \right] \\ &= \int_b^B \hat{L}(\xi; z, b, B) g(\xi) d\xi, \end{aligned}$$

where $\hat{L}(\cdot; x, b, B)$ is the *expected discounted local time* function. It is a weighting function—like a density—for each state ξ up to the stopping time $T(b) \wedge T(B)$, conditional on the initial state x. For the second and third terms it is useful to define

$$\psi(x, b, B) \equiv \mathrm{E}_x \left[e^{-rT} \mid X(T) = b \right] \Pr_x \left[X(T) = b \right],$$

$$\Psi(x, b, B) \equiv \mathrm{E}_x \left[e^{-rT} \mid X(T) = B \right] \Pr_x \left[X(T) = B \right].$$

Thus, $\psi(x, b, B)$ is the expected discounted value, conditional on the initial state x, of an indicator function for the event of reaching the lower threshold b before the upper threshold B is reached. The value $\Psi(x, b, B)$ has a similar interpretation, with the roles of the thresholds reversed. For any $r \geq 0$, clearly ψ and Ψ are bounded, taking values on the interval $[0, 1]$. With w, ψ, and Ψ so defined, the Principle of Optimality implies that v satisfies the Bellman equation

$$\begin{aligned} v(x) = \sup_{b, B, S} \big\{ & w(x, b, B) + \psi\,(x, b, B)\,[v(S) - c] \\ & + \Psi\,(x, b, B)\,[v(S) - c] \big\}, \end{aligned} \tag{1.1}$$

where the optimization is over the choice of the threshold values b and B and return point S.

If X is a Brownian motion or geometric Brownian motion, closed-form expressions can be derived for the functions $\hat{L}$, ψ, and Ψ, and (1.1) provides a direct method for characterizing the optimal policy. If X is a more general diffusion, closed forms for $\hat{L}$, ψ, and Ψ are not available, but fairly sharp characterizations can often be obtained. In either case, several properties of the solution are worth noting. First, it is immediate from (1.1) that the optimal return point S^* maximizes $v(S)$ and does not depend on x. Second, the Principal of Optimality implies that the thresholds b^* and B^* do not depend on x. That is, if b^*, B^* attain the maximum in (1.1) for any $x \in \left(b^*, B^*\right)$, then they attain the maximum for all $x \in \left(b^*, B^*\right)$. The rational decision maker does not alter his choice of thresholds as the state variable evolves. Finally, it is immediate from (1.1) that the value function is a constant function outside the inaction region, $v(x) = v(S^*) - c$, for all $x \notin \left(b^*, B^*\right)$.

An alternative method for characterizing the optimum involves an indirect approach based on (1.1). Notice that if x lies inside the open interval $\left(b^*, B^*\right)$, then for Δt sufficiently small the probability of reaching either threshold can be made arbitrarily small. Hence it follows from (1.1) and the definition of v that

$$v(x) \approx g(x)\Delta t + \frac{1}{1 + r\Delta t} \mathrm{E}_x \left[v(x + \Delta X) \right],$$

where ΔX is the (random) increment to the state over Δt. As will be shown more formally in Chapter 3, if X is a Brownian motion with parameters (μ, σ^2), then a second-order Taylor series approximation gives

$$\mathrm{E}_x \left[v(x + \Delta X) \right] \approx v(x) + v'(x)\mu \, \Delta t + \tfrac{1}{2} v''(x)\sigma^2 \, \Delta t.$$

Using this fact, rearranging terms, and letting $\Delta t \to 0$ gives

$$rv(x) = g(x) + \mu v'(x) + \tfrac{1}{2}\sigma^2 v''(x). \tag{1.2}$$

This equation, a second-order ordinary differential equation (ODE), is called the *Hamilton-Jacobi-Bellman (HJB) equation*. The optimal value function v satisfies this equation on the inaction region, the interval $\left(b^*, B^*\right)$. To complete the solution of a second-order ODE two boundary conditions are needed. In addition, for this problem the thresholds b^* and B^* must be determined. Recall that v is known outside the inaction region, $v(x) = v(S) - c$, all $x \notin \left(b^*, B^*\right)$. The two boundary conditions for (1.2) and the thresholds b^* and B^* are determined by requiring that v and v' be continuous at b^* and B^*. These conditions, called *value matching* and *smooth pasting*, reproduce the solution obtained by maximizing in (1.1).

This approach can be applied to a variety of problems. For inventory or investment models it is natural to assume that there are proportional costs of adjustment as well as fixed costs, and that both types of costs can be different for upward and downward adjustments. Even with these changes, however, the overall structure of the solution is similar to the one for the menu cost model. The main difference is that there are two return points, $s^* < S^*$, where the former applies for upward adjustments, from b^*, and the latter for downward adjustments, from B^*.

Models with two state variables can sometimes be formulated in terms of a ratio so that they have this form as well. An example is a model of investment. Suppose that demand X follows a geometric Brownian motion. Suppose further that labor and raw materials can be continuously and costlessly adjusted, but that capital investment entails a fixed cost. Let revenue net of operating costs for labor and raw materials be $\Pi(X, K)$, and assume that Π displays constant returns to scale. Assume that the fixed cost of adjusting the capital stock, λK, is proportional to the size of the installed base. This cost can be interpreted as the time of managers and the disruption to current production. Assume the proportional cost $P\left(K' - K\right)$ is the price of investment goods multiplied by the size of the investment. Thus, $\lambda K/P$ can also be interpreted as the fraction of the existing capital stock that must be scrapped when new capital is installed.

Under either interpretation, the assumption that demand is a geometric Brownian motion and that Π is homogeneous of degree one together imply that the problem can be formulated in an intensive form that has only one state variable, the ratio $x = X/K$. The optimal policy for the problem so written then takes the form described above. There are thresholds b^*, B^* that define an inaction region and one or two return points inside that region. When the ratio x reaches either boundary of the inaction region, or if the initial condition lies outside that region, the firm immediately invests or disinvests. The return points are the same or different for upward and downward adjustments as the proportional costs are the same or different. Note that in this setting the investment decision involves, implicitly, taking into account the option to invest in the future. Thus, the rule of thumb "invest if the expected discounted returns exceed the cost of investment" does not hold. Instead the problem is one of judiciously choosing when and how much to invest, mindful that investing immediately in effect destroys the opportunity to invest in the near future.

Models with fixed adjustment costs for the state variables can also accommodate control variables that are continuously and costlessly adjustable. These controls affect the evolution of the state variable(s) and may affect the current return as well. An example is a model of portfolio choice and housing purchases. The goal in this model is to examine the effect of home ownership on other parts of the consumer's portfolio. Suppose that the consumer has total wealth Q and has a house of value K. Wealth grows stochastically, with a mean and variance that depend on the portfolio—the mix of safe and risky assets—held by the consumer. Suppose that when a house is sold the owner receives only the fraction $1 - \lambda$ of its value. The fraction λ can be thought of as representing agents' fees, time spent searching, moving costs, and so on. Thus, the fixed cost λK is proportional to the level of the state variable, as in the investment model described above. Assume that the consumer's preferences are homogeneous of some degree. That is, $U(K) = K^\theta/\theta$, where $\theta < 1$. Then the value function is homogeneous of degree θ, and the optimal policy functions are homogeneous of degree one. That is, the optimal policy for purchasing a new house involves only the ratio of total wealth to housing wealth, $q = Q/K$, and has the same form as the policy for the menu cost model.

The new element here is the portfolio, which can be continuously and costlessly adjusted. That is, the consumer can allow her mix of safe and risky assets to depend on her ratio q of total wealth to housing wealth. The question is whether her tolerance for risk varies with this ratio. For example, is her risk tolerance different when the ratio q is near a threshold, so that an adjustment in the near future is likely, and when it is near the return point, as it is just after a transaction? The key technical point is that her portfolio decisions affect the evolution of her wealth, so the state no longer follows an exogenously specified stochastic process between adjustments. Nondurable consumption can also be incorporated into the model as another control that can be continuously adjusted. In this case preferences must be assumed to have a form that preserves the required homogeneity property, but no other restrictions are needed.

In all of the examples considered so far the fixed cost is discrete, and the decision maker adjusts the state variable under his control by a discrete amount if he pays that cost. That is, the fixed cost is lumpy and so are the adjustments made by the decision maker. Models of this type are sometimes called *impulse control* models.

The last section of the book treats a somewhat different class of problems, called *instantaneous control* problems, in which adjustments are continuous. The decision maker chooses a rate of upward or downward adjustment in the state variable and pays a (flow) cost that depends on the rate of adjustment. The cost can have both proportional and convex components, and they can differ for upward and downward adjustments. Optimal policies in these settings share an important feature with the previous class of models, in the sense that there are typically upper and lower thresholds that define an inaction region. When the state is inside this region no adjustments are made. When the state reaches or exceeds either threshold control is exercised, with the optimal rate of adjustment depending on how far the state exceeds the threshold.

Instantaneous control models have an inaction region if the proportional costs are different for upward and downward control. For example, if the purchase price P for investment goods exceeds the sale price p, then investment followed quickly by disinvestment costs $P - p > 0$ per unit. An optimal policy involves avoiding adjustments of this type, creating an inaction region.

Instantaneous control models do not produce the lumpy adjustment that is characteristic of impulse control models, but their implications are similar if control is aggregated over discrete intervals of time. Once the state variable reaches or exceeds the threshold in a model with instantaneous control, it is likely to remain in that vicinity for some time. Consequently control is exercised for some time. Similarly, once the state is well inside the inaction region, there is likely to be a substantial interval of time during which no control is exercised. Hence when aggregated over discrete time intervals total control looks somewhat lumpy, with periods of substantial control and periods of little or no control.

An example is the following inventory problem. The state variable Z is the size of the stock, which in the absence of intervention is a Brownian motion. The manager's objective is to minimize total expected discounted costs. These costs have two components. First there is a (flow) holding cost $h(Z)$ that depends on the size of the stock. The function h is assumed to be continuous and U-shaped, with a minimum at zero. Negative stocks are interpreted as back orders. In addition there is a price $P > 0$ per unit for adding to the stock and a price $p \gtreqless 0$ per unit for reducing it. If p is positive it is interpreted as the revenue per unit from selling off inventory, and otherwise it is interpreted as the unit cost of disposal. In this example there are no other costs of control. Clearly $p \leq P$ is required, so the system is not a money pump, and $p \neq P$ is required to avoid the trivial solution of keeping the stock identically equal to zero. The interest rate $r > 0$ is constant.

Suppose the manager chooses thresholds b, B, and adopts the policy of keeping the inventory inside the closed interval $[b, B]$. To do this, when the state is $Z = b$ he makes purchases that are just sufficient to keep the stock from falling below b, and when the state is $Z = B$ he sells off just enough inventory to keep the state from rising above B.

Let $v(z)$ denote the expected discounted total cost from following an optimal policy, given the initial state z. As before it is useful to think of $v(z)$ as the sum of three parts:

$$\begin{aligned} v(z) = &\text{ expected holding costs} \\ &+ \text{expected cost of control at } b \\ &+ \text{expected cost of control at } B. \end{aligned}$$

Notice that each of these three terms is the expected value of an integral over the entire time horizon $t \in [0, +\infty)$. As before, the key to mathematical tractability is the fact that each of these terms can be written in a convenient way. In particular, the manager's problem can be written as

$$\begin{aligned} v(z) = \min_{b,B} \Bigg[&\int_b^B \pi(\zeta; z, b, B) h(\zeta) d\zeta \\ &+ \alpha(z, b, B)P - \beta(z, b, B)p \Bigg], \quad z \in [b, B], \end{aligned} \tag{1.3}$$

where $\pi(\zeta; z, b, B)$ is the expected discounted local time at each level $\zeta \in (b, B)$, given the initial state z, and

$$\begin{aligned} \alpha\,(z, b, B) &\equiv \mathrm{E}_z \left[\int_0^\infty e^{-rs} dL \right], \\ \beta\,(z, b, B) &\equiv \mathrm{E}_z \left[\int_0^\infty e^{-rs} dU \right], \end{aligned}$$

represent the expected discounted control exercised at the two thresholds. If the underlying process is a Brownian motion explicit formulas are available for α, β, and π, and (1.3) provides a direct method for characterizing the optimal thresholds.

As before, there is also an indirect method for characterizing the optimum. Indeed, the first part of the argument is the same as before: if Z lies inside the interval (b, B), then over a sufficiently short period of time Δt the probability of reaching either threshold is negligible. Using the same second-order Taylor series approximation as before then produces the HJB equation, which in this case is

$$v(z) = h(z) + \mu v'(z) + \tfrac{1}{2}\sigma^2 v''(z).$$

As before, solving the HJB equation requires two boundary conditions, and in addition the optimal thresholds b^* and B^* must be determined. In the present case value matching holds automatically, so requiring v to be continuous at b^* and B^* provides no additional restrictions. Here the two constants and two thresholds required for the solution are determined by requiring that v' and v'' also be continuous at b^* and B^*. These conditions—smooth pasting and *super contact*—reproduce the solution obtained by maximizing (1.3).

Another example of this type of model is an investment problem in which demand $X(t)$ follows a geometric Brownian motion and investment is irreversible. Specifically, suppose that the unit cost of new investment is constant, $P > 0$, but capital has no scrap value, $p = 0$. Suppose, further, that there are no fixed costs and no other adjustment costs, and that the profit flow per unit of capital depends on the ratio of the capital stock to demand, $k = K/X$. The optimal policy in this setting involves choosing a critical value κ for the ratio k. When the ratio k falls to κ, the firm invests just enough to keep the ratio from falling below that threshold. When k exceeds κ, the firm does nothing. If the initial state k_0 is less than κ, the firm makes a one-time purchase of capital sufficient to bring the ratio up to the threshold. It is interesting to compare the optimal policy in this setting with optimal investment in a frictionless world, where the price P applies to sales of capital as well as to purchases. In the frictionless world the optimal policy involves maintaining the ratio k at a fixed level k^f. It can be shown that irreversibility makes the firm less willing to invest, in the sense that $\kappa < k^f$. Larger increases in demand are required to trigger investment, since subsequent reductions in demand cannot be accommodated by selling off installed capital.

The discussion so far has concerned problems faced by individual decision makers, firms, or households. For many questions, however, it is aggregates that are of interest. For example, to assess the role of sticky prices in generating short-run effects from monetary policy, the behavior of the aggregate price level must be determined, as well as the distribution of relative prices across firms. Similarly, to assess the role of hiring and firing rules, the impact on aggregate job creation and destruction is needed. And to assess the role of investment in propagating cyclical shocks, the effect on aggregate investment of a macroeconomic disturbance—such as a change in foreign demand—must be determined.

The difficulty of describing aggregates in models where individual agents face fixed adjustment costs depends largely on the nature of the exogenous shocks. Specifically, it depends on whether the shocks are idiosyncratic or aggregate. If the shocks that agents face are idiosyncratic and can be modeled as independently and identically distributed (i.i.d.) across agents, describing aggregates—at least in the long run—is fairly easy. In these cases the law of large numbers implies that the joint distribution of the shock and the endogenous state across agents converges to a stationary distribution in the long run. This stationary distribution, which also describes long-run averages for any individual agent, is easy to calculate. These distributions are described in Chapters 7, 8, and 10. Thus, describing aggregates is straightforward when the shocks are idiosyncratic.

If agents face an aggregate shock the situation is much more complicated. In settings of this type the law of large numbers is not helpful, and no general method is available for describing aggregates. The main issue is that the distribution of the endogenous states across agents varies over time, and at any point in time it depends on the history of realizations of the shock process. Thus, in general the cross-sectional distribution of the endogenous state does not converge.

For models of this sort there are two possible strategies. The first is to look for special assumptions that permit a stationary distribution. For example, for the menu cost model a uniform distribution of prices is compatible with individual adjustment behavior under certain assumptions about the money supply process. Consequently under these assumptions the model is tractable analytically. Alternatively, one can use computational methods that describe the evolution of the entire distribution. This approach is broadly applicable and it has been pursued successfully in several contexts. These applications are noted as they arise, and the references in these sections describe in more detail the methods they employ.

The goal in this book has been to keep the mathematical prerequisites at a level that is comfortable for builders of economic models. The discussion assumes some background in probability theory and stochastic processes, but an extensive knowledge of these areas is not required. Recursive arguments are used throughout, and familiarity with Bellman equations in discrete time is useful but not necessary.

The rest of the book is organized as follows. Chapters 2–5 introduce some basic mathematical tools. The goal is to provide enough background to permit a fairly rigorous treatment of the optimization problems under study, while keeping the entry barriers low. Thus, the coverage is deliberately selective and proofs are omitted if they play no role later. Chapter 2 introduces stochastic processes, focusing on continuous-time processes. Brownian motions and more general diffusions are defined, as well as stopping times. Chapter 3 treats stochastic integrals, Ito's lemma, occupancy measure, and local time, concepts that are used extensively later. Martingales are discussed in Chapter 4, and the optional stopping theorem is stated. Chapter 5 draws on this material to study the functions ψ, Ψ, and $\hat{L}$ defined above. Explicit formulas are derived for the case in which the underlying stochastic process is a Brownian motion or geometric Brownian motion, and sharp characterizations are provided for more general diffusions. Later chapters make repeated use of the explicit formulas developed in this chapter.

Chapters 6–9 treat a sequence of impulse control problems, displaying a range of applications—price adjustment, investment, and durable goods problems—as well as various modeling devices. These chapters provide a rigorous treatment of the value matching and smooth pasting conditions that describe optima, showing precisely the optimization problems that lead to those conditions.

Chapters 10 and 11 treat instantaneous control models. In Chapter 10 the notion of a regulated Brownian motion, the basis for these models, is introduced and is used to analyze a classic inventory problem. In Chapter 11 a variety of investment models are studied using similar techniques. Optimal control in these settings involves a super contact condition.

Chapter 12 treats two variations on an aggregate menu cost model. In one of them, an aggregate state variable is a regulated Brownian motion. On the substantive side, these models provide useful insights about the role of sticky prices as a source of short-run monetary non-neutrality. On the methodological side, they illustrate how specific assumptions can be used to make aggregate models with fixed costs analytically tractable.

Notes

Bils and Klenow (2004) look at detailed data on the frequency of price changes over 1995–97, and Klenow and Kryvtsov (2008) describe the magnitude of these changes over 1988–2003. Both use data from the U.S. Bureau of Labor Statistics. Doms and Dunne (1998) look at investment at the plant level, and Davis, Haltiwanger, and Schuh (1996) look at job creation and destruction at the plant level. Both use data from the U.S. Bureau of the Census on manufacturing plants for the period 1973–88. See Schachter and Kuenzi (2002) for data on the duration of residence. Vissing-Jorgensen (2002) and the references there discuss evidence on household portfolio behavior.

Many of the ideas developed in this book have appeared elsewhere, at various levels of mathematical rigor and with various emphases on techniques and economic questions. Dixit (1993) is an good introduction to impulse and instantaneous control models, with many examples and illustrations. Dixit and Pindyck (1994) treat a variety of investment problems, with excellent discussions of the economic issues these models address and the empirical predictions they lead to. Harrison (1985) has a detailed treatment of mathematical issues related to instantaneous control models, as well as many examples.

The methods developed here have many applications in finance. Excellent treatments of these problems are available elsewhere, however, as in Duffie (1988, 1996), so they have been neglected here.

The term *Principle of Optimality* was introduced by Richard Bellman. The idea was developed in Bellman (1957), which is still a rich source of applications and problems. See Stokey and Lucas (1989) for a discussion of recursive techniques in discrete time.