Real Options and q Theory

Kerry E. Back

DOI:10.1093/acprof:oso/9780190241148.003.0020

Abstract and Keywords

The theory of perpetual options and dynamic programming are applied to analyze the optimal capital investment of a firm. When investment is continuous and capital is the numeraire, the marginal value of capital is called marginal q. The optimal investment rate is a function of marginal q. When investment is irreversible and there is no depreciation, the optimal time to make each marginal investment is given by the theory of perpetual options. The optimal investment times can also be calculated by dynamic programming. Fluctuations in marginal q add risk to a firm, compared to reversible investment. The Berk-Green-Naik model is an example of a model that relates risk and expected return to size and book-to-market by endogenizing investment.

Keywords: real options, marginal q, q theory, adjustment costs, irreversible investment, dynamic programming, Berk-Green-Naik model

Contents

20.1 An Indivisible Investment Project 515
20.2 q Theory 518
20.3 Irreversible Investment as a Series of Real Options 524
20.4 Dynamic Programming for Irreversible Investment 530
20.5 Irreversible Investment and Perfect Competition 535
20.6 Berk-Green-Naik Model 541
20.7 Notes and References 546
The purpose of this chapter is to show how to apply methods developed earlier—dynamic programming and option pricing—to study optimal corporate investment. Some corporate investment problems are most easily solved by dynamic programming and some are most easily solved by option pricing. Sections 20.3 and 20.4 solve a single problem by the two different methods to show that they give the same result when both are applicable and to illustrate the relationship between them.

Corporate investment is a central topic in corporate finance and has also been of increasing interest in asset pricing research in recent years, as researchers attempt to endogenize firm characteristics, firm risks, and expected stock returns so as to study the cross-sectional relations between characteristics and returns in theoretical models with optimizing and rational firms and investors. The interest in such models stems largely from empirical findings of cross-sectional anomalies such as size, value, and momentum (Section 6.6). One possible explanation of the empirical results is that firm characteristics such as market capitalization and book-to-market ratio are related to priced risk factors that have been omitted from the asset pricing models being tested. Another possible explanation is that errors in estimating betas with respect to risk factors leave explanatory power for firm characteristics. Market value is always related to risk; hence, as Berk (1995) observes, variables based on market value—such as size or book-to-market—can be expected to have explanatory power for returns whenever risk factors are omitted or betas misestimated. To determine whether these explanations for anomalies are plausible, we need a model in which firm characteristics and risks are jointly and endogenously determined. Another motivation for such models is the desire to understand average returns preceding and following corporate financing decisions, for example equity issues or dividend changes. Such decisions are naturally related to investment decisions. For example, the empirical observation that returns are lower on average following seasoned equity issues may be due to the fact that investment converts a growth option into assets in place, lowering risk (Carlson, Fisher, and Giammarino, 2006).
The basic prescription for optimal investment is to invest up to the point that the marginal cost of investment equals the marginal value of capital. The marginal value of capital divided by the price of capital is called marginal $q$. The marginal cost of investment includes both the price of capital and any adjustment costs. Throughout the chapter, we take capital as the numeraire, so its price is always 1. If there are no adjustment costs, then the optimal investment rule is to invest up to the point that marginal $q$ equals 1. If adjustment costs are quadratic, then the optimal investment rate is an affine function of marginal $q$. These and related results are presented in Section 20.2, using the method of dynamic programming.

The simplest application of option pricing theory to corporate investment is to determine the optimal time for a company to make an investment, when the investment is indivisible and irreversible. This is the topic of Section 20.1. If there is no fixed horizon for making the investment, then the investment option is a perpetual call option, and, under certain assumptions, the optimal exercise time is given by the results of Section 19.1. The essential lesson from this model is that if a firm can delay investment—that is, if it has an investment option that is not yet at its maturity—then the simple net present value (NPV) rule of investing if the present value of project cash flows exceeds the project cost is not the correct rule. Instead, the firm should wait until the option is sufficiently far in the money, and how far it should be in the money is given by the formula of Section 19.1. Projects may include abandonment options, which can be analyzed as perpetual puts, and expansion options, which are additional calls. The method described in Section 20.1 can easily accommodate these variations.

Continuous (perfectly divisible) corporate investment can sometimes be analyzed by considering each marginal unit of capital as a separate investment option and applying option pricing theory. This is illustrated in Section 20.3, which uses option pricing theory to solve an optimal investment problem when there is no depreciation of capital and investment is irreversible. The same problem is solved by dynamic programming in Section 20.4. Irreversible investment is (p. 515) a particular type of adjustment cost. When investment is irreversible, then the optimal investment rule is to invest only...
when marginal $q$ equals 1 and to do nothing when marginal $q$ is less than 1.

Section 20.5 analyzes a perfectly competitive industry with irreversible investment and free entry. Because of free entry, growth options are exercised as soon as they reach the money. Therefore, growth options have no value, and the value of each firm is the value of its assets in place. The equilibrium industry output price is a reflected geometric Brownian motion. The equilibrium industry investment process is the singular process that causes the price to be reflected. Section 20.4 explains singular processes, and Section 20.5 explains reflected geometric Brownian motion. An interesting feature of the model is that the risk of a company’s stock is small when the output price is high, because, at the reflection point, increases in demand are absorbed by increases in supply rather than by price increases. Below the reflection point, demand shocks produce price shocks and corresponding shocks to stock returns. In fact, risk is a decreasing function of the output price and consequently an increasing function of the book-to-market ratio: The model predicts that value (high book-to-market) industries have higher risks and higher expected returns than do growth (low book-to-market) industries.

Section 20.6 presents the model of Berk, Green, and Naik (1999), which is a seminal contribution to the literature that endogenizes firm characteristics and risks to determine whether empirical relationships between characteristics and returns are consistent with optimal investment and rational asset pricing. The model is a discrete-time model in which a new investment opportunity appears to each firm each period. The different opportunities have different risks, and the risk of a firm at a point in time depends on the risks of the projects it has taken. It also depends on the values of the projects taken (which determines the value of assets in place) versus the value of growth options (stemming from future investment opportunities). As a consequence, in the cross section of firms, risks and hence expected returns are related to market capitalizations and to book-to-market ratios.

Assume there is no debt and no taxes, so the value of the firm is the value of equity. Assume that markets are complete, so there is a unique SDF process. As discussed in Section 19.3, this implies that all shareholders agree that the firm should
maximize the market value of equity computed according to the unique SDF process. For simplicity, assume also that there is a constant risk-free rate.

20.1 An Indivisible Investment Project
In this section, we consider a single option to make a discrete investment in a project. Let $K$ denote the investment cost, and let $Y_t$ denote the cash flow that the project would generate at date $t$ if the project were adopted prior to $t$. As in Section 19.3, use a version of the Gordon growth model for $Y$ and its value. Assume the following:

- There is a constant risk-free rate $r$.
- $dY_t / Y_t = \mu dt + \sigma dB$ for constants $\mu$ and $\sigma$ and a Brownian motion $B$ under the physical probability.
- The unique SDF process $M$ is such that $e^{\mu t}M_t$ is a martingale.
- $(dY_t / Y_t)(dM_t / M_t) = -\sigma \lambda dt$ for a constant $\lambda$.
- $\mu < r + \sigma \lambda$.
- There are no bubbles.

Set $\delta = r + \sigma \lambda - \mu$. The value of the project’s future cash flows at any date $t$ (if the investment has been made by $t$) is $X_t$, defined as $Y_t / \delta$ (Exercise 13.2). Furthermore, the value $X$ satisfies

$$
\frac{dX}{X} = (r - \delta) dt + \sigma dB^*,
$$

(20.1)

where $B^*$ is a Brownian motion under the risk-neutral probability associated with the SDF process $M$ (Exercise 15.2).

Optimal Exercise and Value
The investment option is a perpetual call option on an asset with price $X$ that has a constant dividend yield $\delta$ (the cash flow stream of the asset is $Y = \delta X$). Therefore, from (19.8), the optimal exercise boundary is

$$
X^* = \frac{\beta K}{\beta - 1},
$$

(20.2)

where $\beta$ is the positive root of the quadratic equation (19.3).

From (19.9), the value of the option, given optimal exercise, is
The Net Present Value Rule

Consider an alternative scenario in which the option can only be exercised at date 0. In this case, the option should be exercised if it is in the money, meaning $X_0 > K$. The NPV of the project in this scenario is $X_0 - K$, and the NPV rule is to (p. 517) invest if the NPV is positive. Clearly it is inappropriate (and costly) to follow this rule if the firm has the flexibility to delay investment. If the firm has the flexibility to delay investment, then it should not invest until

$$X_t \geq \frac{K}{p} - K + \frac{K}{p - 1},$$

as shown in (20.2). Firms usually have the flexibility to delay investment, so the NPV rule is usually inappropriate.

The NPV rule can be modified so that it applies correctly to perpetual options. The key is to recognize that adopting the project imposes two separate costs. One is the exercise price $K$. The other is an opportunity cost: taking the project extinguishes the investment option. The correct investment rule can be described as follows: invest when the adjusted NPV is greater than or equal to zero, where the adjusted NPV is defined as $X_t - K - V$, with $V$ denoting the value of the option. The adjusted NPV is never strictly positive. It is usually negative, because the value of the option is usually greater than the option’s intrinsic value (that is, $V > X_t - K$). At the optimal exercise time, we have value matching ($V_x = X_t - K$), which means that the adjusted NPV is zero. Thus, investing when the adjusted NPV is greater than or equal to zero produces the correct decision. This perspective about the opportunity cost of investing is useful for understanding marginal $q$ in the irreversible investment model discussed in Sections 20.3–20.5.

An Abandonment Option

Now suppose the firm has an option to abandon the project after it is taken. Suppose that abandoning the project produces a positive cash flow $L$ (if the cash flow $L$ were negative, then the project would never be abandoned, given our assumption that the project cash flow $y$ is always positive). The abandonment option is a put option, giving the firm the
right to sell the asset with value $X$ in exchange for receiving $L$. From (19.11), the value of the put option, when optimally exercised, is

$$\left(\frac{1}{1+r}\right)^{\gamma} \left(\frac{X}{T}\right)^{-\gamma},$$

where $\gamma$ is the absolute value of the negative root of the quadratic equation (19.3). Let $f(X)$ denote the project value including the abandonment option; that is,

$$f(X) = X + \left(\frac{1}{1+r}\right)^{\gamma} \left(\frac{X}{T}\right)^{-\gamma}.$$

To determine when the project with the abandonment option should be taken, consider exercising the call option on the project when $X$ first rises to a boundary (p.518) $x^*$. The value at exercise is the intrinsic value $f(x^*) - K$, and the value of receiving $1$ at the exercise time is $(X_t/x^*)^\delta$, so the value of the call option for $X_t < x^*$ with this exercise policy is

$$f(x^*) - KX_t^{\delta} = \left[ x^* - K + \left(\frac{1}{1+r}\right)^{\gamma} \left(\frac{X}{T}\right)^{-\gamma} \right]X_t^{\delta}.$$

where $\beta$ is the positive root of the quadratic equation (19.3). Maximizing this in $x^*$ produces the optimal exercise boundary for the investment option when the project includes the abandonment option. Other project features, such as expansion options, can be addressed similarly by defining $f$ to equal the value of the project including the additional features.

20.2 $q$ Theory

This section derives the optimal investment rule for a firm that can make continuous investments. The rule is expressed in terms of marginal $q$. The market value of the firm depends on its capital stock $K$ and an exogenous stochastic process $X$. Let $J(K, X)$ denote the market value of the firm at date $t$. Use subscripts to denote partial derivatives. Marginal $q$ is defined to be $J_t(K, X)$, which is the marginal value of capital. Average $q$ is the average value of capital, which is $J(K, X)/K_t$. Average $q$ is often proxied empirically by the sum of the market values of debt and equity divided by the sum of their book values. The investment rule is in terms of marginal $q$, but, under circumstances described below, marginal $q$ equals average $q$.  

The basic model is as follows. A firm has a given initial capital stock $K_0$. It chooses an investment rate $I_t$ at each date $t$, and its capital stock $K_t$ evolves as

$$\dot{K}_t = I_t - \rho K_t,$$

(20.4)

where the dot notation denotes the derivative with respect to time. The constant $\rho$ is the depreciation rate. The firm’s operating cash flow depends on an exogenous stochastic process $X$ and is given by a function $\pi(K_t, X_t)$. Assume $X$ is a Markov process relative to the risk-neutral probability with dynamics

$$dX = \mu(X)dt + \sigma(X)dB^*$$

(20.5) \hspace{1cm} \text{(p.519)}

for functions $\mu$ and $\sigma$, where $B^*$ is a Brownian motion relative to the risk-neutral probability. For example, we could have $\mu(x) = \mu x$ and $\sigma(x) = \sigma x$ for constants $\mu$ and $\sigma$, in which case $X$ is a geometric Brownian motion relative to the risk-neutral probability. Let $E^*$ denote expectation with respect to the risk-neutral probability.

The firm’s net cash flow is its operating cash flow minus the cost of investment. The cost of investment is given by a function $\theta(K_t, I_t)$. The cost of investment includes the price of capital plus any adjustment costs. Examples are given below.\(^2\)

The market value of the firm is the maximum over investment policies of the risk-neutral expectation of discounted cash flows:

$$J(K_0, X_0) \overset{\text{def}}{=} \max_{I} E^* \int_0^\infty e^{-r_t} \pi(K_t, X_t) - \theta(K_t, I_t) dt.$$  

(20.6)
**HJB Equation and the First-Order Condition**

The HJB equation is

\[ 0 = \max_{Q(k, x)} \left[ \pi(k, x) - \theta(k, x) - rJ(k, x) + \frac{1}{2} J''(k, x) \right] \]

This is based on the principles explained in Section 14.5, using the dynamics (20.4) of \( \kappa \), the dynamics (20.5) of \( x \), and the fact that there is discounting in (20.6) at rate \( r \). Note that there are no second-order terms in the HJB equation with respect to the \( k \) argument, because \( d\kappa \) is of order \( dt \). Assume \( \theta \) is differentiable in the investment rate. Then, the first-order condition for the maximum in the HJB equation is that \( \theta = J' \). Thus, the optimal investment rate is the rate that equates the marginal cost of investment \( \theta \) to the marginal value of capital \( J' \) (which is marginal \( q \)).

**Examples of Investment Costs**

Most of the examples we will discuss have the property that \( \theta \) is linearly homogeneous in \( (k, \dot{x}) \). In that case, it is convenient to define a function \( \phi \) by \( \phi(y) = \theta(1, y) \). We then have

\[ \theta(k, \dot{x}) = k\phi(1/k) = k\phi(i/k) \]

This implies that \( \theta(k, \dot{x}) = \phi(i/k) \). Therefore, the first-order condition for the maximum in the HJB equation is \( \phi(i/k) = J'(k, x) \).

An example of a linearly homogeneous \( \theta \) is \( \theta(k, \dot{x}) = ai + bi^2/k \) for \( a, b > 0 \). This quadratic case is quite natural. The linear term can be interpreted as the price of capital and the quadratic term as an adjustment cost. As remarked before, we take capital to be the numeraire. This means that we take \( a = 1 \) and consider quadratic cost functions \( \theta(k, \dot{x}) = i + bi^2/k \) for some \( b \). Note that if \( i < 0 \) then the “cost” \( i \) is actually a cash inflow, representing the receipts from selling capital. The quadratic term implies that it is more costly to install capital quickly than it is to install capital slowly. For example, installing at rate \( i = 1 \) for a unit period of time costs \( 1 + b/k \), but installing at rate \( i = 2 \) for half of a period costs \( 1 + 2b/k \). The extra cost could represent higher prices paid to suppliers for quicker delivery, overtime costs for employees, and so on. Notice that if \( i < 0 \) then the quadratic term implies that receipts fall below \( i \). This could represent higher labor costs for dismantling and moving capital or depressed prices for capital as a result of a “fire sale.”
Any smooth strictly convex function $\theta$ can be interpreted similarly to the quadratic function. We can represent such a $\theta$ by a Taylor series expansion, and the first-order term $\theta(0, k)$ can be interpreted as the price of capital. Strict convexity implies that the second-order term $\theta''(0, k)$ is positive, like the coefficient $b > 0$ in the quadratic case.

A linear $\theta$ represents costless adjustment. In this case, capital can be bought and sold at a fixed price. Costless adjustment is unrealistic but nevertheless interesting as a benchmark. With $\theta(k, \cdot) = i$, the first-order condition in the HJB equation states that $J_k = 1$. This is an equation that can be solved for $k$ as a function of $\chi$. Thus, with costless adjustment, capital will be adjusted instantaneously in response to changes in $\chi$ in such a way as to maintain marginal $q$ equal to 1. In contrast, with quadratic investment costs $\theta(k, \cdot) = i + bk^2/k$, the first-order condition implies that marginal $q$ satisfies $J_k = 1 + bi/k > 1$ when the firm is investing ($i > 0$) and satisfies $J_k = 1 + bi/k < 1$ when the firm is disinvesting ($i < 0$). Capital is not adjusted instantly to make marginal $q$ equal to 1, because of the adjustment costs.

Another interesting special case is $\theta(k, \cdot) = i - ci$, where $c < 1$ and where $i^* = \max(i, 0)$ and $i = \max(-i, 0)$. In this case, $c$ is the resale price of capital when the firm disinvests. This function is linearly homogeneous. It is also convex, but it is not strictly convex, and it is not differentiable at $i = 0$. In this case, the first-order condition for the maximization in the HJB equation is that $J_k = 1$ when the firm is investing, $J_k = c$ when the firm disinvests, and, when $c < J_k < 1$, the firm neither invests nor disinvests.

A special case of the preceding is $\theta(k, \cdot) = i^*$. In this case, there is no market for the firm’s used capital. If the capital produces nonnegative operating cash flows, then the firm will never dispose of capital at a zero price, so, for practical purposes, investment is irreversible.

Another possibility not yet considered is that there could be fixed costs of investment. Fixed costs are costs that are independent of the scale of investment. For example, we could have $\theta(i, k) = i^* + F_1(i^*0)$ for a constant $F$. In this case, the resale price of capital is zero, and there is a fixed cost $F$ to be paid.
whenever investment is positive. This cost function is not linearly homogeneous and is not studied further in this chapter.

Main Results of \( q \) Theory

Assume the investment cost function \( \vartheta(k, \eta) \) is linearly homogeneous, and define \( \phi(y) = \vartheta(1, y) \). Recall that marginal \( q \) is defined as \( J_k \).

(i) If \( \phi \) is differentiable and strictly convex, then the optimal investment-to-capital ratio is a strictly increasing function of marginal \( q \). Specifically, \( \frac{i}{k} = (\phi)^{-1}(J_k) \).

(ii) In the quadratic case—meaning \( \vartheta(k, \eta) = i + b\eta^2/k \) for a constant \( b > 0 \)—the optimal investment-to-capital ratio is an affine function of marginal \( q \). Specifically, \( \frac{i}{k} = (J_k - 1)/(2b) \).

(iii) If \( \pi(k, x) \) is linear in \( k \), then marginal \( q \) equals average \( q \); that is, \( J_q(k, x) = J(k, x)/k \).

As is discussed below, operating cash flows are linear in capital when there is perfect competition and constant returns to scale in production. Together, (ii) and (iii) imply that with linear operating cash flows and quadratic investment costs, the optimal investment-to-capital ratio is an affine function of average \( q \). That relationship has been tested extensively.

The assumption that \( \phi \) is differentiable and strictly convex excludes the cost function \( \vartheta(i, k) = r - c_l \). We discuss that case later in this section and again with \( c = 0 \) in Sections 20.3–20.5.

To prove (i) and (ii), we use the first-order condition \( J_q(k, x) = \phi(i/k) \) from the HJB equation. Strict convexity of \( \phi \) implies that \( \phi \) is strictly increasing, hence invertible. Therefore, \( \frac{i}{k} = (\phi)^{-1}(J_k) \). When \( \vartheta(k, \eta) = i + b\eta^2/k \), we have \( \phi(y) = y + by^2 \), and \( \phi(y) = 1 + 2by \). Hence, \( (\phi)^{-1}(z) = (\frac{z - 1}{2b}) \), and the first-order condition implies \( \frac{i}{k} = (J_k - 1)/(2b) \).
To establish (iii), we show that the market value of the firm is linear in \( k \); that is, \( J(k, x) = kq(x) \) for some function \( q \). Our assumptions are that \( n(k, x) \) is linear in \( k \), meaning that \( n(k, x) = f(x) \) for some function \( f \), and that \( \theta \) is linearly homogeneous. Given those assumptions, the date–0 market value of net cash flows for any investment process \( I \) is

\[
E \int_0^\infty e^{-\gamma t} [f(X_t) - \phi(I_t | K_t)] dt = K_0 \cdot E \int_0^\infty e^{-\gamma t} \frac{K_t}{K_0} [f(X_t) - \phi(I_t | K_t)] dt. 
\]

Make the change of variables \( Y_t = K_t / K_0 \) and \( Z_t = I_t / K_0 \). Then, the date–0 market value of the firm is

\[
J(K_0, X_0) = K_0 \max_{I} \mathbb{E} \int_0^\infty e^{-\gamma t} [f(Y_t) - \phi(Z_t | Y_t)] dt. 
\]

(20.7)

Here, the maximization takes as given that \( Y_0 = 1 \) and that the dynamics of \( Y \) are \( \dot{Y}_t = Z_t - \rho Y_t \). The value of the maximization problem in (20.7) is independent of \( K_0 \). Denote it as \( q(X_0) \). Then, (20.7) states that \( J(K_0, X_0) = K_0 q(X_0) \). Because \( K_0 \) and \( X_0 \) are arbitrary, we have \( J(k, x) = kq(x) \) for all \( k \) and \( x \). Thus, \( J_k(k, x) = q(x) = J(k, x) / k \).

Resale Price of Capital, Marginal \( q \), and Inactivity

It is very reasonable to assume that the firm can sell capital only at a discount to the price it pays for capital. This motivates the investment cost function \( \theta(k, i) = i' - ci' \) with \( c < 1 \) that is discussed above. It can be generalized to include adjustment costs, as, for example, \( \theta(k, i) = i' - ci' + bi'^2 / k \). This cost function is strictly convex but not differentiable at zero, due to the kink at the origin caused by the lower resale price of capital. Consequently, the optimal investment rate is not a strictly increasing function of marginal \( q \). With the cost function \( \theta(k, i) = i' - ci' + bi'^2 / k \), the optimal investment rate is

\[
i = \begin{cases} 
\frac{I_k - 1}{I_k - c} & \text{if } I_k > 1, \\
0 & \text{if } c < I_k < 1, \\
\frac{I_k - c}{I_k - c} & \text{if } I_k \leq c.
\end{cases} 
\]

Thus, the investment rate is insensitive to marginal \( q \) when marginal \( q \) is greater than the resale price of capital but below the price of capital. This region of inactivity weakens the correlation between marginal \( q \) and the investment rate. This issue is discussed further in Section 20.3.

Examples of Operating Cash Flows
A common model is to assume that the firm has a Cobb-Douglas production function in capital and labor: \( y = Ak^{\alpha}l^{\beta} \) with \( A, \alpha, \psi > 0 \) and \( \alpha + \psi \leq 1 \). Labor \( l \) is hired in a perfectly competitive market. Suppose first that the firm is a price taker in the output market. Then, given the output price \( p \), wage rate \( w \), and capital stock \( k \), the firm chooses labor \( l \) to maximize its operating cash flow \( Apk^{\alpha}l^{\beta} - wl \). Solving for the optimal \( l \) and substituting, the firm’s operating cash flow is \( \pi(k, x) = xk^\eta \), where \( x \) is a function of the output price, the wage rate, and the productivity factor \( A \). Setting \( \eta = \alpha/(1 - \psi) \), the operating cash flow is \( \pi(k, x) = xk^\eta \). If there are decreasing returns to scale, meaning \( \alpha + \psi < 1 \), then \( \eta < 1 \) and \( \pi \) is strictly concave in \( k \). If there are constant returns to scale, meaning \( \alpha + \psi = 1 \), then \( \eta = 1 \) and \( \pi \) is linear in \( k \). In this model, the exogenous stochastic process \( X \) is the price or the wage rate or the productivity factor \( A \), or a combination of the three.

Alternatively, assume the firm is a monopolist in the output market, and assume the industry demand curve satisfies \( p = by^{-1/\varepsilon} \) for a constant \( \varepsilon > 1 \) (\( \varepsilon \) is the elasticity of demand). Maximizing over the labor input as before, we can compute the operating cash flow as \( xk^\hat{\eta} \), where \( x \) is a function of the demand parameter \( b \), the wage rate \( w \), and the productivity factor \( A \), and where \( \eta = \hat{\alpha}/(1 - \hat{\psi}) \) with \( \hat{\alpha} = \alpha(\varepsilon - 1)/\varepsilon \) and \( \hat{\psi} = \psi(\varepsilon - 1)/\varepsilon \). In this case, \( \eta < 1 \) and the operating cash flow is strictly concave in \( k \) even if there are constant returns to scale in production.

In the Cobb-Douglas model, marginal \( q \) is less than average \( q \) if there are decreasing returns to scale or if the firm is a monopolist. Marginal \( q \) equals average \( q \) in the Cobb-Douglas model if and only if there are constant returns to scale and perfect competition.

It is also possible for marginal \( q \) to be greater than average \( q \). An example is \( \pi(k, x) = xk - c \) for a constant \( c \). In this example, \( c \) represents fixed costs that are independent of the capital stock. Such costs are called operating leverage and could be due to overhead expenses in the form of executive salaries, for example.

The Investment Return in Discrete Time
Assuming $\theta$ is differentiable, the ratio $J_k / \theta_i$ is the marginal return on investment spending. To see this, note that the marginal unit of investment generates a cost equal to $\theta_i$. So, spending $\theta_i$ creates a unit of investment, and spending 1 creates $1/\theta_i$ units of investment. Each unit of investment creates a unit of capital, which contributes $J_k$ to the value of the firm. Therefore, spending 1 contributes $J_k / \theta_i$ to the value of the firm. The ratio $J_k / \theta_i$ is therefore the marginal return on investment spending, or, more simply, the investment return.

The first-order condition in the continuous-time model is that the investment return is equal to 1. This means that exactly one dollar of value is created by the marginal dollar spent. In discrete time models, it is natural to assume there is a delay between the time of investment spending and the time that the spending affects the stock of capital: investment at $t$ affects capital at $t+1$. In this case, the value of the marginal dollar spent at $t$ is not known until $t+1$, because the value of capital at $t+1$ depends on $X_{t+1}$. Consequently, the first-order condition for optimal investment is a bit different.

Assume capital evolves as $K_{t+1} = (1-\rho)K_t + I_t$. At each date $t$, the firm chooses investment $I_t$ to maximize

$$E_t \left[ \frac{M_{t+1}}{M_t} J(1-\rho)K_t + I_t X_{t+1} \right] - \theta(K,I_t)$$

The first term is the date-$t$ market value of the cum-dividend value of the firm at $t+1$; that is, it is the date-$t$ market value of owning the firm at $t+1$ prior to distributions of cash flows at $t+1$. Assuming we can interchange differentiation and expectation, the first-order condition for the optimization can be arranged as

$$E_t \left[ \frac{M_{t+1}}{M_t} \frac{J(K_t + X_{t+1})}{\theta(K,I_t)} \right] = 1.$$

(20.8)

Thus, in a discrete-time model, the first-order condition is that the investment return be priced like an asset return. Consequently, investment returns can be used instead of, or in conjunction with, asset returns to identify the SDF process.

20.3 Irreversible Investment as a Series of Real Options
This section studies the q theory model of Section 20.2 with the investment cost function $\theta(k, q) = i$. This means that the resale price of capital is zero, and there are no adjustment costs, except for the fact that downward adjustments occur at the zero price. In the model of this section, the marginal value of capital is always positive, so the firm will never sell capital at a zero price. Thus, investment is effectively irreversible. In this section, we analyze irreversible investment using the real options approach developed in Section 20.1. We analyze the same model using dynamic programming in Section 20.4.

The investment cost function $\theta(k, q) = i$ is linearly homogeneous, but it is not differentiable at $q = 0$, and it is not strictly convex. Thus, the q theory described in Section 20.2 does not directly apply. However, the model with $\theta(k, q) = i$ (p.525) is quite tractable, and we can even solve it in closed form. We will solve it in this section by considering each marginal unit of capital as a separate investment option. We will see that the firm’s marginal q equals 1 at the optimal exercise times for the investment options and is less than 1 at all other times. Thus, exercising the investment options optimally can be described thus: Invest when marginal q equals 1.

Assume there is no depreciation of capital ($\rho = 0$). Depreciation and other extensions of the model can be analyzed by dynamic programming as discussed in Section 20.4.

Operating Cash Flows
Assume the operating cash flow at each date $t$ is $Y_t k_t^\eta$, where $\eta < 1$, and $Y$ satisfies the Gordon growth model assumptions in Section 20.1. Then, for fixed capital $k$, the value at any date $t$ of all future operating cash flows $Y_t k_t^\eta$ is $X_t k_t^\eta$, where $X = Y / \delta$ is a geometric Brownian motion under the risk-neutral probability with dynamics (20.1). Write the operating cash flow function as $\pi(k, \lambda) = \delta x k^\eta$. As discussed in Section 20.2, the condition $\eta < 1$ follows from Cobb-Douglas production and monopoly power in the product market, with either constant or decreasing returns to scale. Perfect competition is discussed in Section 20.5.

Capital Stock Process and the Firm’s Objective
Denote the initial capital stock of the firm by $k_0$. The investment cost function $\theta(k, q) = r^q$ means that there are no adjustment costs for positive investment (beyond the fixed price of capital). Therefore, it could be optimal to make a discrete investment in an instant of time. Such discrete investments represent jumps in the capital stock process. Let $K_t$ denote the firm’s capital stock at each date $t$, including any possible discrete investment at time $t$. If there is a discrete investment at date 0, then $K_0 > k_0$. There will be a discrete investment at date 0 if the given initial capital stock is too small. However, it is not optimal to make discrete investments at any other times.

In order to accommodate jumps in the capital stock process, we do not assume that the investment rate is well defined, and we drop the evolution equation $K_t = I_t - rK_t$. The issue with the investment rate being undefined actually goes beyond jumps, as will be explained. Under our assumptions, the paths of the capital stock process are nondecreasing (investment is irreversible and there is no depreciation) and the firm chooses the capital stock process to maximize

$$E \int_0^\infty e^{-r[t(\zeta_t, X_t)]} dt - dK_t$$

(20.9)

(p.526) Here, the integral $\int_0^\infty e^{-\pi t} dK_t$ adds up each increment to capital discounted back to date 0 at rate $r$. These are the discounted investment costs. Formally, the integral is a Lebesgue-Stieltjes integral.

**Optimal Exercise and Value of Investment Options**

The firm has a continuum of investment options, one for each infinitesimal unit of capital. At a capital stock level of $k$, the cash flow produced by the marginal unit of investment is the marginal operating cash flow $\pi(k, x) = \eta x / k^{1-\eta}$. The market value of receiving $\delta x$ is $\chi$, so the market value of the marginal cash flow stream at date $t$ is $\eta \chi / k^{1-\eta}$.

Consider a perpetual call option with a strike equal to 1 on an asset with market value $\eta \chi / k^{1-\eta}$. The market value inherits the dynamics of $\chi$; thus, it is a geometric Brownian motion with drift $r - \delta$ under the risk-neutral probability. Because the strike is equal to 1, it follows from (19.8) that the call should be
exercised when the underlying value \( \eta x_t / k^{1-\eta} \) reaches \( \beta / (\beta - 1) \), or equivalently, when \( x_t \) reaches

\[
\frac{\beta k^{1-\eta}}{(\beta - 1)\eta},
\]

(20.10)

where \( \beta \) is the positive root of the quadratic equation (19.3).

From (19.9), the value of the option prior to exercise is

\[
(\beta - 1)^{\beta - \frac{\eta x_t}{\beta k^{1-\eta}}}.
\]

(20.11)

Assets in Place and Growth Options

The value of the firm at any date \( t \) is the value of the cash flow stream produced by the firm’s existing capital \( K_t \) plus the sum of the values of the infinitesimal investment options. The value of the cash flow stream produced by the firm’s existing capital \( K_t \) is called the value of assets in place. The cash flow at \( u > t \) from capital \( K_t \) is \( \delta x_t K_t^\eta \), so the value of assets in place at date \( t \) is \( x_t K_t^\eta \). The sum of the values of the infinitesimal investment options is called the value of growth options. It equals the integral of (20.11) over all capital stock levels greater than \( K_t \).

This is

\[
\int_{k_t}^{\infty} (\beta - 1)^{\beta - \frac{\eta x_t}{\beta k^{1-\eta}}} \, dk = (\beta - 1)^{\beta - \frac{\eta x_t}{\beta}} \frac{1}{\beta(1-\eta)-1} \beta k^{1-\eta}.
\]

(20.12)

(\text{p.527}) Actually, (20.12) is correct only when \( \beta(1-\eta) > 1 \).

Otherwise, the left-hand side of (20.12)—the value of growth options—is infinite. Given that \( \eta < 1 \), the condition \( \beta(1-\eta) > 1 \) holds if and only if \( \beta > 1/(1-\eta) \). Using the quadratic formula to calculate \( \beta \) as the positive root of the quadratic equation (19.3) and employing some algebra, it can be seen that the condition \( \beta > 1/(1-\eta) \) is equivalent to

\[
\frac{\eta r^2}{2(1-\eta)^2} + \frac{r - \delta}{1-\eta} < r.
\]

(20.13)

Assume this condition holds for the remainder of this section.\(^3\)

Optimal Capital Stock Process

From (20.10), the optimal investment policy is to invest whenever \( x_t \) reaches
This means that the capital stock process \( K \) should be the smallest nondecreasing process such that, for all \( t \),

\[
\frac{\beta K_t^{1-q}}{(\beta - 1)\pi_t}.
\]

(20.14)

This is equivalent to

\[
X_t \leq \frac{\beta K_t^{1-q}}{(\beta - 1)\pi_t}.
\]

This is equivalent to

\[
K_t \geq \left(\frac{\beta - 1}{\beta}\right)^{\frac{1}{q}} X_t.
\]

(20.15)

Of course, we also must have \( K_t \geq k_0 \) for all \( t \), where \( k_0 \) is the given capital stock at date 0. The smallest nondecreasing process satisfying these conditions for all \( t \) is given by

\[
K_t = \max\left(k_0, \max_{s \leq t} \left(\frac{\beta - 1}{\beta} X_s\right)^{\frac{1}{q}}\right).
\]

(20.16)

This is the optimal capital stock process. According to this formula,

\[
k_0 < \left(\frac{\beta - 1}{\beta}\right)^{\frac{1}{q}} \Rightarrow \left(\frac{\beta - 1}{\beta}\right)^{\frac{1}{q}} X_t > k_0.
\]

Thus, there is a discrete investment at date 0 when \( k_0 \) is smaller than is optimal at date 0.

The paths of \( K \) are continuous (except for the possible jump from \( k_0 \) to \( K_0 \) at date 0) and nondecreasing, but their derivatives with respect to time are zero at almost all times \( t \) (that is, except for times in a set having zero Lebesgue measure) with probability 1. This is a consequence of the properties of Brownian motion paths. In particular, the paths of \( K \) are not equal to the integrals of their derivatives! We could regard the investment rate as being zero almost all of the time but infinite on the zero Lebesgue measure set of times at which investment occurs. However, it is safer to simply regard the investment rate as being undefined. The paths of \( K \) are called singular with respect to the Lebesgue measure, and the process \( K \) is called a singular process. An example of a path of \( K \) is shown in Figure 20.1.

Marginal q
The value of the firm is the value $X_t K_t^q$ of assets in place plus the value (20.12) of growth options. Differentiate the sum with respect to capital $k$ to obtain

$$J_t(k, x) = \frac{nx}{k^{1-\gamma}} - (\beta - 1)^{\gamma-1} \left( \frac{nx}{\beta k^{1-\gamma}} \right)^{\beta}.$$  

(20.17)

This is marginal $q$ (because capital is the numeraire). The investment boundary defined by (20.10), namely,

$$\{(k, x) \mid x = \frac{\beta k^{1-\gamma}}{(\beta - 1)^{\gamma}}\},$$  

(20.18)

is the level curve $\{(k, x) \mid J_t(k, x) = 1\}$. Furthermore, $J_t < 1$ for $x$ less than the boundary (20.18). Therefore, investing at the boundary is equivalent to (p.529) investing when marginal $q$ is equal to 1. The boundary is plotted in Figure 20.2 in Section 20.4.

To understand the formula (20.17) for marginal $q$, note that the first term is the value of the future cash flows contributed by the marginal unit of capital. To understand the second term, it is useful to recall the previous discussion of adjusted NPV. There are two consequences of investment: the first is that investment generates additional future cash flows; the second is that investment extinguishes

**Figure 20.1 Optimal capital stock.**

The top panel presents a path of a geometric Brownian motion $X$. The middle panel plots $f(X_t) = [(\beta - 1)\eta X_t / (\beta)^{(1-\eta)}$. The dotted line in the middle panel is $K_t = \max(k_t, \max_{sq} f(X_t))$, which is the optimal capital process (20.16). The bottom panel...
the option to invest. The second term in (20.17) reflects the value of the option that is extinguished by investing the marginal unit of capital, which (p.530) is the option value (20.11). Thus, marginal $q$ is the difference between the value of the marginal cash flows and the value of the option. According to the adjusted NPV rule, the firm should invest when $\chi - V \geq K$, where $\chi - V$ is the value of future cash flows minus the value of the option. In the current setting, the value of future cash flows minus the value of the option is marginal $q (J_k)$, and the strike is the price of capital (1). Therefore, the adjusted NPV rule is to invest when $J_k \geq 1$. Adjusted NPV is never strictly positive, and it is never the case that $J_k > 1$. We always have $J_k \leq 1$ and investment is optimal only when $J_k = 1$. This is illustrated in Figure 20.1.

Hysteresis and the Correlation between $q$ and Investment

Hysteresis means path dependence. The optimal capital stock process in this model is dependent on the history of the exogenous process $\chi$, not just on the current value of $\chi$.

Consider Figure 20.1 and time 0.5. By time 0.5, the process $\chi$ has dipped from its previous high. Consequently, the firm has too much capital at time 0.5, compared to the optimum at that time. If the firm could choose its capital stock freely—understanding that it would be irreversible going forward—then it would choose the amount denoted as $f(X_0)$ in the figure, which is the right-hand side of (20.15). At that level, marginal $q$ would equal 1, whereas it is actually much lower than 1 at time 0.5. Hysteresis implies that we cannot predict the firm’s capital stock at any point in time based on the firm’s circumstances at that time. We have to know how the firm arrived at where it is. This occurs in any model with costly adjustment of capital.

Hysteresis is due to costly adjustment. In this model, in which the investment cost function is not differentiable at zero, another phenomenon also occurs. As Figure 20.1 illustrates, there are many fluctuations in marginal $q$ that do not produce changes in investment, because $q$ remains below 1. For example, the uptick in marginal $q$ between times 0.5 and 0.6...
does not lead to any increase in investment. This implies that empirical regressions of investment on proxies for marginal $q$ or for changes in marginal $q$ may produce weaker results than we might otherwise expect.

20.4 Dynamic Programming for Irreversible Investment
Dynamic programming gives the same result as the option pricing method applied in the previous section. The option pricing method is perhaps simpler, but dynamic programming is more general. To highlight the equivalence of the two methods, we use dynamic programming to solve the model of the previous section.

**HJB Equation and First-Order Condition**
The method of dynamic programming for this model is basically the same as for portfolio choice. The only new element is that the control process $K$ is singular. Itô’s formula still allows us to calculate $dJ(K, X_t)$ even when $K$ is singular. Because $K$ is continuous and has finite total variation, it has zero quadratic variation. Consequently, Itô’s formula and the rules for multiplying differentials apply just as if $dK$ were “something $dt$.” Specifically,

$$dJ = J_t dK + J_t dX + \frac{1}{2} J_t \sigma^2 dX^2.$$ 

Therefore, the HJB equation is

$$\max_{dK \geq 0} \left\{ \pi dt - dK - rJ dt + J_t dK + J_t (r - \delta) x dt + \frac{1}{2} J_t \sigma^2 x^2 dt \right\} = 0.$$ 

(20.19)
The maximization in the HJB equation is a constrained maximization, because, heuristically, $dK \geq 0$ as a result of $K$ being nondecreasing (irreversible investment). The first-order condition is that $J_k \leq 1$ and that $(J_k - 1) dK = 0$. The latter condition (called the complementary slackness condition in the Kuhn-Tucker theory of constrained optimization) states that the capital stock increases only when marginal $q$ is equal to 1.5

**Fundamental ODE**
The first-order condition implies that the $dK$ terms cancel in the objective function in the HJB equation (20.19). Therefore, the HJB equation implies

$$\pi - rJ + J_t (r - \delta) x + \frac{1}{2} J_t \sigma^2 x^2 = 0.$$ 

(20.20)
Even though $J$ is a function of $k$ and $x$, the derivatives in this equation are only with respect to $x$. It is an ODE for a function $x \mapsto J(k, x)$ that should hold for (p.532) each fixed $k$ (for a range of $x$ values depending on $k$ that is described as the inaction region below). As such, it is a nonhomogeneous version of the fundamental ODE (19.2). It states that, for a fixed capital stock $k$, the expected return of the asset paying dividends $\pi$ and with value $J(k, x)$ is equal to the risk-free rate under the risk-neutral probability.

Depreciation

For this paragraph, let $I_t$ denote cumulative investment through date $t$ (instead of the rate of investment at date $t$ as in Section 20.2). Without depreciation, $K_t = k_0 + I_t$. We could include depreciation by specifying instead that $dK = -\rho K dt + dl$ for a constant $\rho$. In this case, the HJB equation would be

$$\max_{q_2} \{r dt - dl - r J dt + J (r - \delta) x dt + \frac{1}{2} J_{xx} \sigma^2 x^2 dt \} = 0.$$ 

The first-order condition is still that investment occur only when $J_k = 1$. The fundamental ODE would be a PDE:

$$\pi - r J - \rho J_k + (r - \delta) x J_x + \frac{1}{2} \sigma^2 x^2 J_{xx} = 0.$$ 

In the remainder of this section, we continue to study the model without depreciation.
Action and Inaction Regions

Figure 20.2 plots the optimal investment boundary (20.18), which we know from the analysis in Section 20.3, though we will calculate it independently in this section. The ODE (20.10) must hold in what is called the inaction region. This is the region in which $\chi$ is below the investment boundary. The region above the investment boundary is called the action region. It is never reached, except possibly at date 0. Because the action region is not reached after date 0, the ODE (20.20) does not need to hold in the action region.

It is simple to calculate the value function in the action region based on the value function in the inaction region. If the firm is in the action region at date 0, then the optimal action is to add capital so as to move immediately to the boundary of the region. Figure 20.2 illustrates an increase from $k = k_0$ to $K_0$ based on the value of $x_0$. The value function at the point $(k, x_0)$ in the action region is the value at $(K_0, x_0)$ minus the cost of moving from $k$ to $K_0$, which is $K_0 - k$. Note that investing to increase capital beyond $K_0$ would be suboptimal, because $J_k < 1$ in (p.533) the inaction region. Thus, we can describe the value function in the action region as

$$J(k, x_0) = J(K_0, x_0) - (K_0 - k),$$

where $K_0$ is such that $(K_0, x_0)$ is on the investment boundary. Given $x_0$, this $K_0$ is the same for all $k < K_0$. Thus, $J$ is linear in $k$ in the action region with coefficient equal to 1. We therefore obtain the following derivatives in the action region: $J_k(k, x) = 1$ and $J_{kk}(k, x) = 0$. 

![Illustration of Action and Inaction Regions](image-url)
Smooth Pasting and Super Contact

To solve the problem using only dynamic programming methods, we need to find the investment boundary and the function \( J \) below the boundary (in the inaction region). We do that by solving the ODE (20.20) with the requirement that the solution paste smoothly at the boundary with the value in the action region described in the previous paragraph. That is, we require

\[
J_\delta(k, x) = 1 \quad \text{and} \quad J_\delta(k, x) = 0
\]

at the boundary. The first of these conditions is commonly called smooth pasting, and the second is called super contact. However, the first condition is value matching and the second is smooth pasting in the option framework. To see that, we can use the formula (20.17) for \( J_k \). Set \( s = \eta x/k^{1-\eta} \). The condition \( J(k, x) = 1 \) is equivalent to

**Figure 20.2** Action region, inaction region, and investment boundary. The curve is the investment boundary (20.18). The area below the curve is the inaction region, and the area above is the action region. If the initial condition \((k_0, x_0)\) is in the action region, then the firm should invest a discrete amount (shown as \( K_0 - k_0 \) in the plot) to move immediately to the investment boundary.

**Figure 20.3** A reflected Brownian motion. The upper path is a path of a Brownian motion \( B \). The dashed line is the running maximum of \( B \). Subtracting the running maximum from the Brownian motion produces a Brownian motion reflected from above at 0.
The left-hand side of (20.21) is the value (20.11) of the infinitesimal investment option. The right-hand side is the value of the underlying asset for the investment option minus its exercise price. Thus, the condition \( J_k(k, x) = 1 \) is equivalent to value matching: the value of the option equals its intrinsic value. Using the same formula for \( J_k \), we can calculate \( J_{kk} \) by using the chain rule as

\[
J_{kk}(k, x) = \frac{\delta}{\delta x} \left\{ (\beta - 1)^{\delta} \left( \frac{x}{\beta} \right) \right\} \cdot \frac{x}{\beta} - \frac{\delta x}{\beta}.
\]

Therefore, the condition \( J_{kk}(k, x) = 0 \) is equivalent to

\[
\frac{\delta}{\delta x} \left\{ (\beta - 1)^{\delta} \left( \frac{x}{\beta} \right) \right\} = 1,
\]

which is smooth pasting for the option value.

Solving the ODE with the Boundary Conditions

The ODE (20.20) is a nonhomogeneous version of the ODE (19.2). As discussed in Exercise 19.4, it can be solved by valuing the nonhomogeneous part as a cash flow to be received forever and adding that to the general solution (19.4) of the homogeneous part. Here, the nonhomogeneous part is \( \pi(k, x) = \delta x k^\delta \). This is the cash flow from assets in place. Its value is the value \( x k^\delta \) of assets in place. Thus, the general solution of the ODE (20.20) is \( x k^\delta + ax^\gamma + bx^\delta \), where \( \gamma \) is the absolute value of the negative root of the quadratic equation (19.3) and \( \beta \) is the positive root. The coefficients \( a \) and \( b \) can depend on \( k \). Because the value converges to zero as \( x \) converges to zero, the coefficient of \( x^\gamma \) must be zero. Hence, it must be that

\[
J(k, x) = x k^\delta + b(k) x^\delta
\]

for a function \( b \). The term \( b(k) x^\delta \) is the value of growth options. The function \( b(k) \) and the investment boundary can be derived from the smooth pasting and super contact conditions \( J_k = 1 \) and \( J_{kk} = 0 \). We leave the details as an exercise. The resulting formula \( b(k) x^\delta \) for the value of growth options is the same as the formula (20.12) obtained in Section 20.3.
This section describes industry equilibrium assuming perfect competition with free entry and irreversible investment. The definition of industry equilibrium is that firms invest optimally, and growth options never get strictly in the money. When they reach the money—that is, when the intrinsic value reaches zero—entry of new firms or expansion of existing firms exhausts the options that are at the money. Therefore, growth options have no value, and the value of each firm is the value of its assets in place. As in the previous section, optimality implies that investment occurs when and only when marginal $q$ is equal to 1. At the end of the section, we explain how fluctuations in $q$ result in risky returns for the firm’s shareholders.

**Depreciation and Investment**

For the remainder of the chapter, let $I_t$ denote cumulative investment through date $t$ (instead of the rate of investment at date $t$ as in Section 20.2). We allow for depreciation. Assume $dK = -\rho K dt + dl$ for a constant $\rho$. Require $I$ to be a nondecreasing process (investment is irreversible). If there is a discrete investment at date 0, then $I_0$ denotes the amount of the investment and $K_0 = k_0 + I_0$; otherwise, $I_0 = 0$.

**Output and Operating Cash Flow**

Assume Cobb-Douglas production with constant returns to scale as discussed in Section 20.2; that is, $y = A K_t^{1-a}$. Take $A$ to be constant and the wage rate $w$ to be constant. From each firm’s point of view, the output price $P$ is exogenous. We leave it as an exercise to show that, at the optimal labor input, output is

\[ c_y P_t^{1-a} K_t \]

(20.22)

and operating cash flow is $c_y P_t^{1-a} K_t$ for constants $c_y$ and $c_n$.

**Industry Output Price**

The output price $P$ depends on supply and demand. Supply is determined by the industry capital stock. Denote the industry capital stock by $K_{ind}$. By summing (20.22) over firms, we see that industry output is

\[ Q_t = c_y P_t^{1-a} K_{ind} \]

(20.23)
Assume the industry demand curve is \( P_t = Z_t Q_t^{-\epsilon} \), where \( Z \) is a geometric Brownian motion under the risk-neutral probability. Demand is higher when \( Z \) is higher. The parameter \( \epsilon \) is the elasticity of demand.

Let \( P_t \) denote the price at which supply equals demand, and define \( Y_t = c_n P_t^{1/\alpha} \). Then, operating cash flow is

\[
c_n P_t^{1/\alpha} K_t = Y_t K_t.
\]

(20.24)

We leave it as an exercise to show that

\[
\frac{dY}{Y} = (r - \delta) dt + \sigma dB - \kappa \left( \frac{d I^{\text{ind}}}{K^{\text{ind}}} \right)
\]

(20.25)

for constants \( \delta, \sigma, \) and \( \kappa \), where \( B^* \) is a Brownian motion under the risk-neutral probability and where \( I^{\text{ind}} \) denotes cumulative industry investment.

**Reflected Geometric Brownian Motion**

We solve for equilibrium in two steps. First, we define \( I^{\text{ind}} \) in such a way that \( Y \) defined by (20.25) is a geometric Brownian motion reflected from above at some constant \( \gamma^* \). This construction has the property that firms in the industry invest \( (I^{\text{ind}} \) increases) if and only if \( Y \) is at \( \gamma^* \). Then, we calculate \( \gamma^* \) by requiring marginal \( q \) to equal 1 if and only if \( Y \) is at \( \gamma^* \). This produces the equilibrium price process \( P = (Y/c_n)^\alpha \) and the equilibrium industry capital stock process.

Before beginning the construction of \( I^{\text{ind}} \), it is useful to explain reflected Brownian motions and reflected geometric Brownian motions. Figure 20.3 illustrates how the subtraction of a singular process can cause a Brownian motion to be reflected. Figure 20.3 shows a path of a Brownian motion that is reflected from above at 0. The reflected Brownian motion is the process \( N \) defined by \( N_t = B_t - \max(B_t | s \leq t) \). The process \( \max(B_t | s \leq t) \) is called the running maximum of the Brownian motion \( B \). Denote it by \( U \), so we have \( N = B - U \). Like the optimal capital stock process in Sections 20.3 and 20.4, the process \( U \) is a singular process—it has continuous and nondecreasing paths, but the derivative of each path with respect to time is zero at almost all times (with probability 1). The reflected Brownian motion evolves as \( dN = dB - dU \). At almost all times,
\[ dU = 0, \] (p.537) and the dynamics of \( N \) are the same as the dynamics of \( B \). However, subtraction of the singular process \( U \) from the Brownian motion \( B \) leads to reflection at 0.

The same construction allows us to reflect a \((\mu, \sigma)\)-Brownian motion from above at 0. If \( Z \) is such a Brownian motion (so \( dZ = \mu dt + \sigma dB \) for a standard Brownian motion \( B \)), then \( N \) defined as \( N_t = Z_t - \max\{Z_s : s \leq t\} \) is a \((\mu, \sigma)\)-Brownian motion reflected from above at 0. Furthermore, we can easily modify the construction to reflect \( Z \) from above at any number \( a \). We simply take \( N_t = Z_t - \max\{Z_s - a : s \leq t\} \). This ensures that

\[ N_t \leq Z_t - (Z_t - a)^+ = a + (Z_t - a) - (Z_t - a)^+ \leq a. \]

Finally, we can construct a reflected geometric Brownian motion as \( e^N \).

(p.538) Industry Investment and Price Processes

Now we define an industry investment process \( I^{\text{ind}} \) that causes \( Y \) to be a geometric Brownian motion reflected from above at \( Y^* \). Here, we consider any \( Y^* \geq Y_0 \). Later, we compute the equilibrium \( Y^* \). Define

\[
Z_t = \log Y_0 + \left( r - \frac{1}{2} \sigma^2 \right) t + \sigma B^*_t \tag{20.26a}
\]

and

\[
U_t = \max_{s \leq t} (Z_s - \log Y^*)^+. \tag{20.26b}
\]

and set \( N = Z - U \). Then, \( N \) is an \((r - \delta - \sigma^2/2, \sigma)\)-Brownian motion under the risk-neutral probability reflected from above at \( \log Y^* \).

Define \( I^{\text{ind}} \) by

\[
dI^{\text{ind}} = \frac{k^{\text{ind}}}{k} dU. \tag{20.26c}
\]

Then, \( Y \) defined as

\[
Y_t = e^{rt} \tag{20.26d}
\]

satisfies (20.25) and is a geometric Brownian motion under the risk-neutral probability reflected from above at \( Y^* \).

Furthermore, \( U \) and \( I^{\text{ind}} \) increase only when \( Y \) is at \( Y^* \).
A reflected geometric Brownian motion is a Markov process. We are taking the risk-free rate to be constant. Hence, the value at date $t$ of a claim to a reflected geometric Brownian motion cash flow process $Y_u$ for $u \geq t$ depends on $Y_t$. We use $X_t$ to denote this value below. The value $X_t$ is a function of the contemporaneous cash flow $Y_t$.

The industry output price is $P = (Y/c_0)^q$. Consequently, it is also a geometric Brownian motion reflected from above at $p^*$ defined as $(y^*/c_0)^q$. The industry investment process (20.26c) is the smallest nondecreasing process that maintains the industry output price $P$ below $p^*$. We will calculate the equilibrium reflection point $y^*$ and hence the equilibrium $p^*$. In equilibrium, firms invest only when the industry output price reaches $p^*$, and they invest just enough to keep the price from ever rising above $p^*$.

**Marginal and Average $q$**

The value of a firm is the value of assets in place plus the value of growth options. Part of the definition of industry equilibrium is that growth options have no value, because investment in the industry occurs whenever growth options reach the money, due to free entry. Thus, the value of each firm is the value of its assets in place. The operating cash flows are $Y_tK_t$, so the value of a firm’s assets in place at date $t$ is $X_tK_t$, where we define

$$X_t = \mathbb{E}\int_t^\infty e^{-\sigma y} Y_u du$$

(20.27)

Because the value of a firm is $XK$, both marginal and average $q$ are equal to $\frac{X}{K}$. We will calculate $X$ assuming $Y$ is a geometric Brownian motion reflected at $y^* \geq Y_0$. Then, we will determine the equilibrium $y^*$ from the requirement that $X_t = 1$ when $Y_t = y^*$.

As remarked before, $X_t$ depends on $Y_t$. For $y \leq y^*$, set

$$f(y) = \mathbb{E}\left[\int_t^\infty e^{-\sigma y} Y_u du | Y_t = y\right]$$

assuming $Y$ is a geometric Brownian motion reflected from above at $y^*$. Then, $f(Y_0)$ is the value at date $t$ of an asset that pays $Y$ as a dividend. Calculating $f$ is similar to but not quite the same as calculating the value of an asset that pays a cash
flow until a hitting time as in Section 19.1. If cash flows end at
the hitting time, then \( f(y^*) = 0 \). However, in the problem here,
the cash flows continue, but risk vanishes at the hitting time,
because of the reflection. Thus, \( f(y^*) = 0 \). We derive this fact in
the following paragraph.

Assume \( \delta > 0 \). Equate the expected rate of return under the
risk-neutral probability of the asset that pays \( Y \) to the risk-free
rate. This yields

\[
Ydr + \mathbb{E}[f(Y)dY + \frac{1}{2}f(Y)dY^2] = rf(Y)dt.
\]

(20.28)

There are \( dt \) terms and \( dY^{\text{red}} \) terms in \( dY \), and they must
separately match on the two sides of (20.28). Equating the
\( dt \) terms, using the dynamics (20.25) of \( Y \), gives

\[
y + (r - \delta)yf(y) + \frac{1}{2}\sigma^2 y^2 f(y) = rf(y).
\]

(20.29)

This is a nonhomogeneous version of the fundamental ODE
(19.2). The only remaining term on either side of (20.28) is the
singular part of \( dY \) in \( f(Y)dY \). It is equal to zero if and only if
\( f(y^*) = 0 \).

Previously, we solved nonhomogeneous versions of the
fundamental ODE by valuing a claim to the nonhomogeneous
part as a cash flow to be received forever. We can do the same
here, ignoring reflection in the valuation. The value of
receiving \( Y \) forever if it is a geometric Brownian motion under
the risk-neutral probability with drift \( r - \delta \) is \( \frac{Y}{\delta} \). Clearly,
\( f(y) = \frac{Y}{\delta} \) satisfies the ODE (20.29). The general solution of the
ODE is \( \frac{Y}{\delta} + ay^\gamma + by^\beta \), where \( y \) is the absolute value of the
negative root and \( \beta \) is the positive root of the quadratic
equation (19.3).

\((\text{p.540})\) The boundary condition \( \lim_{y \to 0} f(y) = 0 \) implies \( a = 0 \), so
\( f(y) = \frac{Y}{\delta} + by^\beta \). The boundary condition \( f(y^*) = 0 \) implies

\[
b = -\frac{1}{\delta Y^\beta - \frac{1}{\beta}}.
\]

(20.30)

Therefore, when there is reflection at \( y^* \), marginal and average
\( q \) are

\[
X_t = \frac{Y_t}{\delta} + bY_t^\beta = \frac{Y_t}{\delta}\left[1 - \frac{1}{\beta}\left(\frac{Y_t}{\delta}\right)^{\beta - 1}\right]
\]
When \( Y_t \) reaches \( y^* \), marginal and average \( q \) (20.30) are equal to

\[
\frac{(\beta - 1)y^*}{\delta \beta}.
\]

**Equilibrium Reflection Point**

In the construction (20.1), there is investment in the industry when and only when \( Y_t = y^* \). We know from general theory that investment is optimal when and only when marginal \( q \) equals 1. Therefore, the construction (20.1) is consistent with industry equilibrium if and only if marginal \( q \) equals 1 when \( Y_t = y^* \).

Equating (20.31) to 1, we see that the equilibrium reflection point is

\[
y^* = \frac{\delta \beta}{\beta - 1}.
\]

The equilibrium output price is \( P = (y/c_0)^r \), so \( P \) is a geometric Brownian motion reflected at

\[
p^* = \left(\frac{\delta \beta}{(\beta - 1)c_0}\right)^r.
\]

**Irreversible Investment, \( q \), and Risk**

The value of a firm is \( J(K_0, X_t) = X_tK_t \). From this fact and the fact that the capital process \( K \) has zero quadratic variation, we obtain

\[
\frac{dJ(K_0, X_t)}{dX_t} = \frac{dK_t}{K_t} - \frac{dX_t}{X_t} - \rho dt + \frac{dI_t}{K_t}.
\]

The rate of return earned by the firm’s shareholders is

\[
\frac{r(K_0, X_t, dI_t, dJ(K_0, X_t))}{X_t} = \frac{X_tK_t}{X_tK_0} - \frac{dX_t}{X_t} - \rho dt + \frac{dI_t}{K_t}.
\]

Investment occurs only when \( X_t = 1 \), and this implies \( J(K_0, X_t) = K_t \). Therefore, the \( dI_t \) terms in the shareholders’ return cancel, and the return is

\[
\frac{Y_tK_t dt}{J(K_0, X_t)} + \frac{dX_t}{X_t} - \rho dt.
\]

If investment were perfectly reversible (\( \theta(k, i) = i \)), then we would always have marginal \( q \) equal to 1, and the return would be the risk-free return in equilibrium. However, with irreversible investment, fluctuations in marginal \( q \) add the risk for the firm’s shareholders.
From the formula (20.30) for \( x \) and the dynamics (20.25) of \( y \), we can calculate that the volatility of \( x \) (and hence the risk of the stock return) is

\[
\frac{Y}{\delta} \left[ 1 - \left( \frac{Y}{y^\alpha} \right)^{\delta - \rho} \right] \sigma.
\]

(20.35)

See Exercise 20.3. The volatility is decreasing in \( Y \), decreasing to zero at \( Y = y^\alpha \). Equivalently, the volatility is decreasing in the industry output price \( P \), decreasing to zero at \( P = p^\alpha \).

Equivalently, the volatility is decreasing in \( x \), decreasing to zero at \( x = 1 \). The book-to-market ratio of a firm in this model is \( K_0 / J(K_0, X_0) = 1 / X_0 \). All firms in the industry have the same book-to-market ratio. If we consider different industries that are ex ante identical, then firms in industries with higher book-to-market ratios (lower values of \( x \)) will have higher risks and higher expected returns. The reason for the correlation between output price (or book-to-market ratio) and risk is that increases in demand are absorbed by increases in supply when the price is at the reflection point. However, below the reflection point, demand shocks produce price shocks, which produce shocks to stock prices.

20.6 Berk-Green-Naik Model

This section presents a model due to Berk, Green, and Naik (1999) of a firm that confronts a sequence of indivisible investment options. The options are for projects with varying risk; consequently, the risk of the firm changes over time. The model is a tractable model in which we can derive relationships between firm risks, expected returns, book-to-market ratios, and size (market capitalization). It is an example of an equilibrium model with rational agents in which book-to-market and size predict returns in the cross section of stocks: As in the data, high book-to-market (value) stocks and small stocks have higher average returns than do low book-to-market (growth) stocks and large stocks.

(p.542) Projects and Capital

Unlike the other models discussed in this chapter, there are no timing options in this model. A project is available only at the date it appears, so the NPV rule determines optimal investment. A single new project arrives at each date, investment in a project can only be made at the date it arrives, and investment in each project is irreversible. The first project
arrives at date 0. If the firm invests in the project, its operating cash flows begin at date 1. Capital in all projects depreciates at a common rate $\delta$. There is a maximum feasible investment $I$ in any project. There are constant returns to scale in operation in each project up to the feasible scale. The capital at date $u$ in a project that arrived at date $t < u$ is $\delta^{t-u}I_u$, where $I_u$ is the investment made at date $t$ (which will be either 0 or $I$).

The Net Present Value Rule

The operating cash flow generated at date $u$ by a project that arrived at $t < u$, per unit investment at $t$, is $\delta^{t-u}C_{u}$ where $C_{u}$ is an exogenous random variable observable at date $u$, and where $\delta$ is the depreciation rate mentioned in the previous paragraph. The NPV of a unit investment in the project that arrives at $t$ is

$$\mathbb{E}\left[\sum_{u=1}^{\infty} \frac{M_u}{M_t} \delta^{u-t}C_u\right] - 1,$$

(20.36)

where $M$ is the SDF process. Define $\chi_t = 1$ if the NPV is nonnegative and $\chi_t = 0$ if the NPV is negative. An optimal investment process is $I_t = \chi_t I$. Assuming optimal investment, the capital stock of the firm at date $t$ is

$$K_t = I_t \sum_{u=0}^{t-1} \delta^{t-u} \chi_u.$$

SDF Process

Assume

$$\log M_{t+1} = \log M_t - r_t - \frac{1}{2} \lambda^2 - \lambda \epsilon_{t+1},$$

(20.37)

for a constant $\lambda$ (the market price of risk) and a stochastic process $r$, where $\epsilon$ is a sequence of independent standard normal random variables. This implies

$$\mathbb{E}\left[\frac{M_{t+1}}{M_t}\right] = e^{r_t},$$

(p.543) so $r_t$ is the continuously compounded risk-free rate from date $t$ to date $t+1$. To obtain precise formulas, some assumption about the distribution of the risk-free rate process must be made. However, for our purposes, the risk-free rate process can be quite general.

Project Cash Flows
Assume for each $s$ that $(C_{x_{s+1}} C_{x_{s+2}} \ldots)$ is an IID lognormal sequence. Specifically, assume

$$\log C_t = \log \hat{C} - \frac{1}{2} \phi^2 + \phi \xi_{x_t},$$

(20.37b)

for $t > s$, where $\hat{C}$ is a constant, $\phi$ is observable at time $s$, and $(\xi_{x_{s+1}} \xi_{x_{s+2}} \ldots)$ is a sequence of independent standard normal random variables. From the usual rule for means of exponentials of normals, we have, for $t \geq s$,

$$E\left[ \frac{M_{t+1}}{M_t} \right] = e^{\gamma_t - \hat{C} \xi_{x_t}},$$

(20.38)

where we set $\gamma_t = \lambda \phi \text{corr}(\xi_{x_{t+1}}, \xi_{x_{s+1}})$ (assume the correlation depends only on information available at $s$). Thus, the value at $t$ of the cash flow $C_{x_{t+1}}$ is its expected value discounted continuously at the risk-adjusted rate $r_t + \gamma_t$. Assume $(\beta_{t+1}, \beta_{t+2}, \ldots)$ is an IID sequence that is independent of the $\epsilon$'s and $\xi$'s. This completes the assumptions of the model.

Value of Assets in Place

By iterated expectations, (20.38) generalizes as follows:7 for $s \leq t < u$,

$$E\left[ \frac{M_t}{M_s} C_{x_u} \right] = e^{\beta_{t+1} \hat{C} P(u)},$$

(20.39)

(p.544) where $P(u)$ denotes the price at $u$ of a discount bond maturing at $u$. Define

$$D_u = \sum_{s=t+1}^u \delta^{s+1} P(u).$$

Note that $D_u$ is the value of a perpetual bond with coupons declining at rate $1 - \delta$. It follows from (20.39) that the value at $t$ of the cash flows produced at $t+1, t+2, \ldots$, by a project that arrived at $s \leq t$ is

$$E\left[ \sum_{s=t+1}^u \frac{M_{s-1}}{M_t} \delta^{s+1} X_s e^{\phi \xi_{x_s}} \right] = \delta^{t+1} X_t e^{\phi \xi_{x_t}} D_t,$$

(20.40)

Set

$$\gamma_t = -\log \left( \sum_{s=t+1}^u \frac{\delta^{s+1} X_s e^{\phi \xi_{x_s}}}{K_{s-1}} \right),$$

(20.41)
This means that $e^\gamma$ is a weighted average of the risks of projects in which the firm has invested by date $t$, with the weight being the fraction $\delta^{s}x_{s}^{}I/K_{t+1}$ of capital invested in the project when operations begin at date $t+1$ (recall that $K_{t+1}$ is known at date $t$, because it depends on investment decisions made at date $t$ and before).

Let $A_{t}^{}$ denote the value at $t$ of the cash flows produced at $t+1$, $t+2$, ..., by all projects that arrived at $t$ or before. This is the value of assets in place at date $t$. From (20.40) and (20.41), we have

$A_{t}^{} = \hat{C}_{t}D_{t}\sum_{s=0}^{1} \delta^{s}x_{s}^{}Ie^{-\beta_{s}^{}s} = e^{\gamma}K_{t+1}\hat{C}_{t}D_{t}$

(20.42)

Note that $K_{t+1}\hat{C}_{t}$ is the expected cash flow at date $t+1$, conditional on date–$t$ information. Thus, the value of assets in place at date $t$ is the expected cash flow at $t+1$ multiplied by a risk-adjusted value of the perpetual bond.

**Value of Growth Options**

Setting $s = t$ in (20.40) gives the value of (20.36). Thus, the firm invests at $t$ ($x_{t}^{} = 1$) if and only if

$e^{\beta_{t}^{}D_{t}} > 1$.

This is true when interest rates are low, implying that the value $D_{t}$ of the perpetual bond is high, or when project risk $\beta_{t}$ is low. The value at $t$ of the option to invest at $t$ is

$I(e^{\beta_{t}^{}\hat{C}_{t}D_{t} - 1})$,

(p.545) and the value at $t$ of the option to invest at $u > t$ is

$I\left[ E_{t}^{u} \left( \frac{M_{u}^{}(e^{\beta_{u}^{}\hat{C}_{u}D_{u}^{} - 1})}{M_{t}^{}(e^{\beta_{t}^{}\hat{C}_{t}D_{t} - 1})} \right) \right]$.

The total value of growth options at $t$ is $G_{t}$ defined by

$G_{t} = I\sum_{u=t}^{\infty} E_{t}^{u} \left( \frac{M_{u}^{}(e^{\beta_{u}^{}\hat{C}_{u}D_{u}^{} - 1})}{M_{t}^{}(e^{\beta_{t}^{}\hat{C}_{t}D_{t} - 1})} \right)$

(20.43)

**Expected Return**

The total value of the firm after the distribution of cash flows at date $t$ is $A_{t} + G_{t}$. Denote this by $S_{t}^{}$. So, we have

$S_{t} = e^{\gamma}K_{t+1}\hat{C}_{t}D_{t} + G_{t}$

(20.44)
The main purpose of this exercise is to compute the expected return and determine how it relates to characteristics of the firm. The conditional expectation at date \( t \) of the operating cash flow at date \( t+1 \) is \( K_{t+1} \hat{C} = \varphi_t A_t / D_t \), and the conditional expectation of investment cash flow is \(- \mathbb{E}[X_{t+1}]\). The expected return from \( t \) to \( t+1 \) for an owner of the firm is therefore

\[
\varphi_t \frac{A_t}{D_t} + \frac{\mathbb{E}[A_{t+1}] - \mathbb{E}[X_{t+1}]}{S_t} + \mathbb{E}[G_{t+1}] = \mathbb{E}[A_{t+1}] - \mathbb{E}[X_{t+1}] + \mathbb{E}[G_{t+1}].
\]

(20.45)

The value at \( t+1 \) of assets in place equals the value at \( t+1 \) of the assets in place at \( t \) plus the value at \( t+1 \) of assets added at \( t+1 \). This decomposition is

\[
A_{t+1} = \delta A_t + \chi_{t+1} \mathbb{E}[\hat{C}D_{t+1}].
\]

It follows that

\[
\mathbb{E}[A_{t+1}] - \mathbb{E}[X_{t+1}] = \delta A_t \mathbb{E}[\hat{C}D_t] + \mathbb{E}[\mathbb{E}[\varphi_{t+1} \hat{C}D_{t+1}] - \eta \chi_{t+1}].
\]

Substituting this into (20.45) shows that the expected return equals

\[
\varphi_t \frac{A_t}{D_t} + \frac{\mathbb{E}[G_{t+1}] - \mathbb{E}[\varphi_{t+1} \hat{C}D_{t+1} - 1]}{S_t}.
\]

(20.46)

The factor multiplying \( A_t/S_t \) is the expected return on assets in place. Note that \((1 + \delta \mathbb{E}[D_{t+1}]) / D_t \) is the expected return on the perpetual bond. The \((p.546)\) factor multiplying \( G_t/S_t \) is the expected return on growth options, including the growth option that matures at \( t+1 \). The sequence of project risk realizations \( \beta_\varphi, \ldots, \beta_t \) and the interest rates faced by the firm through date \( t \) determine the firm risk \( \gamma_t \) and the relative importance of assets in place versus growth options. In conjunction with the interest rate environment at date \( t \), these firm characteristics determine the expected return.

**Expected Returns, Book-to-Market, and Size**

Given a sample of ex-ante identical firms in this model, the firms will be distinguished at date \( t \) by their risks \( \gamma_t \) and capital stocks \( K_{t+1} \) (because these depend on the project risk realizations \( \beta_\varphi, \ldots, \beta_t \)). The firm risk \( \gamma_t \) is not directly observable.

We can rewrite the expected return by substituting \( K_{t+1} \hat{C} \) in
place of $e^{r}A_{t}/D_{t}$ as the expected operating cash flow and substituting $A_{t}=S_{t}-G_{t}$. Make these substitutions in (20.46) to obtain

$$\frac{K_{t+1}C}{S_{t}}+\delta E_{t}\left[\frac{D_{t+1}}{D_{t}}\right] - \frac{E_{t}\left[\gamma_{t+1}[\gamma_{t}C_{t}D_{t}+1]-1\right]}{S_{t}} - \delta G_{t}[D_{t+1}/D_{t}].$$

The interesting feature of this formula is that the first term is proportional to the book-to-market ratio $K_{t+1}/S_{t}$, the second term depends only on the distribution of interest rates, and the numerator of the third term depends only on the distribution of interest rates, so the third term is inversely proportional to the market value $S_{t}$. Thus, given a sample of ex-ante identical firms, expected returns will vary across the sample depending on book-to-market and size.

20.7 Notes and References

Tobin (1969) uses the symbol $q$ to denote the market value of a firm divided by the replacement cost of its capital. This is now called average $q$. The $q$ theory in Section 20.2 is due to Hayashi (1982) and Abel (1985). Formula (20.8) for the investment return in discrete time is due to Cochrane (1991).

Important early work on real options includes Brennan and Schwartz (1985), McDonald and Siegel (1986), and Dixit and Pindyck (1994). Many of the issues discussed in this chapter are covered by Dixit and Pindyck. The indivisible project model of Section 20.1 has been widely applied and is covered in textbooks such as Trigeorgis (1996).

The effect of asymmetric information on the optimal exercise time in the indivisible project model of Section 20.1 is studied by Morellec and Schürhoff (2011), Grenadier and Malenko (2011), and Bustamante (2012). Morellec and Schürhoff and Bustamante assume the project is financed by raising equity, and firms wish to be viewed as having high project quality because that increases the price at which shares can be issued, minimizing the dilution of existing shareholders. They show that firms with good projects will generally exercise investment options earlier in order to signal project quality, though Bustamante shows that all managers will exercise early (there is a pooling equilibrium) in “hot markets.” Grenadier and Malenko (2011) assume that managers wish to maximize their own utility rather than
shareholder value and with a specific type of utility function show that managers with good projects will sometimes delay investment.

The monopoly model presented in Sections 20.3 and 20.4 is a special case of the model studied by Abel and Eberly (1996). Cooper (2006) solves a version of that model in which there is also a fixed cost to invest and relates the risks and hence expected returns of firms to their book-to-market ratios. The model of perfect competition with irreversible investment studied in Section 20.5 is due to Leahy (1993). The discussion of irreversibility, marginal $q$, and risk at the end of Section 20.5 is based on Kogan (2004). Kogan analyzes a model in which investment is a singular process, as in Section 20.5, and he also analyzes a model in which there is a maximum feasible investment rate. In that model, marginal $q$ can exceed 1—when it is above 1, firms invest at the maximum rate. In that model, risk increases as $q$ decreases below 1, as in Section 20.5, and it also increases as $q$ rises above 1.

For oligopoly versions of the irreversible investment model, see Baldursson (1998), Grenadier (2002), Aguerrevere (2009), Back and Paulsen (2009), and Steg (2012). In the oligopoly model, firms play a continuous-time stochastic game with singular strategies. The theory of such games has so far been developed only for open-loop strategies, meaning that the investment of each firm depends on the anticipated investment of other firms (via the Nash equilibrium condition) but does not depend on the realized investment of other firms. Open-loop equilibria are not subgame perfect. The difference between open-loop and closed-loop equilibria in deterministic investment games is explained in the textbook by Fudenberg and Tirole (1992).

The singular control problem of monopoly investment studied in Sections 20.3 and 20.4—in which there is no depreciation—is an example of what is called a monotone follower problem in the optimal control literature. The relation between the dynamic programming method and optimal stopping (which is optimal exercise in the option context) for monotone follower problems is developed by Karatzas and Shreve (1984). A more practitioner-oriented exposition of the equivalence between
option pricing and dynamic programming (or decision tree analysis) is given by Smith and Nau (1995).

(p.548) To apply the condition “invest when marginal $q$ equals 1” in the irreversible investment model, we first have to calculate marginal $q$, which is the partial derivative $J_k$ of the value function $J$. An equivalent condition that does not require prior calculation of the value function is given by Bank (2005) for the case in which there is no depreciation. Bank shows that, given a capital stock process $K$, we can define a stochastic process $D$ with the property that

$$D_t = E \int_t^\infty e^{-\eta_r(K_s, X_s)} dr - e^{-\eta}$$

for all stopping times $t$. Given some technical conditions, necessary and sufficient conditions for $K$ to be optimal are that (i) $D \leq 0$ and (ii) $\int_0^\infty D_t dK_t = 0$. These are versions of the statements that (i) marginal $q$ never exceeds 1, and (ii) investment occurs only when marginal $q$ equals 1. Steg (2012) uses Bank’s result to analyze the oligopoly model with irreversible investment.

The theory of singular control is relevant for issues in finance other than corporate investment. For example, it applies to portfolio choice with proportional transactions costs (Davis and Norman, 1990) and has been applied to international finance (Dumas, 1992). Textbook treatments of singular control include Harrison (1985, 2013), Øksendal and Sulem (2007), and Stokey (2009).

Gomes, Kogan, and Zhang (2003) solve a model similar to that of Berk, Green, and Naik (1999) while endogenizing the SDF process by assuming a representative investor with CRRA preferences. In their model, the conditional CAPM holds, but empirical tests of the CAPM on simulated samples show that size and book-to-market have additional explanatory power for average returns, due to misestimation of time-varying betas. In the Gomes-Kogan-Zhang model, projects differ by productivity (rather than by covariance with an exogenously specified SDF process as in the Berk-Green-Naik model). However, the Gomes-Kogan-Zhang model shares the feature of the Berk-Green-Naik model that all firms have equal growth options. Thus, in the Gomes-Kogan-Zhang model, a firm with more productive projects is what is generally defined to be a growth firm because it has a lower book-to-market ratio, yet growth
options constitute a lower fraction of its value. The model generates a value premium, because growth options are a greater fraction of the value of a high book-to-market firm and growth options are riskier than assets in place. However, this implies that value firms have higher cash flow durations than growth firms, which is inconsistent with the data (see Section C.4 of Zhang, 2005).

Zhang (2005) analyzes industry equilibrium in a perfectly competitive model with an exogenously specified SDF process, assuming asymmetric quadratic adjustment costs (with a higher cost for disinvesting than for investing) and fixed operating costs. The SDF process has a countercyclical market price of risk. The adjustment costs produce higher risk for firms with excess capital, which (p.549) generally are firms with higher book-to-market ratios. In combination with the countercyclical market price of risk, this produces a value premium. Li, Livdan, and Zhang (2009) and Livdan, Saprina, and Zhang (2009) extend the Zhang model by incorporating equity issues, dividend changes, and capital structure choice.

Carlson, Fisher, and Giammarino (2004) solve a monopoly model with irreversible investment in which there is only a discrete set \( K_0 < K_1 < K_2 \) of feasible capital stocks. They incorporate quasi-fixed operating costs (operating costs that depend on the capital stock). This “operating leverage” produces additional risk for assets in place beyond the additional risk already due to irreversibility and thereby generates a value premium. Carlson, Fisher, and Giammarino (2006) extend the model to analyze risk and expected returns preceding and following equity issuance. Kogan and Papanikolaou (2012) survey investment models and their implications for firm characteristics and expected returns.

Exercises

20.1. Consider the optimization problem

\[
\max_l \quad pAk^{\alpha l^\alpha} - wl
\]

that arises in the perfect competition model of Section 20.5. Compute constants \( c_0 \) and \( c_n \) such that, at the optimal \( l \),

\[
Ak^{\alpha l^\alpha} = c_0 p^{l^\alpha/l^k},
\]

\[
pAk^{\alpha l^\alpha} - wl = c_n p^{l^\alpha/l^k}.
\]

20.2. As in Section 20.5, assume
\[ \frac{dZ}{Z} = \mu_t \, dt + \sigma_t \, dB_t \]

for constants \( \mu_t \) and \( \sigma_t \) and a Brownian motion \( B_t \) under the risk-neutral probability. Match industry supply (20.23) to industry demand \( Q = (Z_t/P_t)^f \) to compute the equilibrium output price \( P_t \).

Define \( Y_t = c_n \pi^\alpha_t \). Compute constants \( \delta_t, \sigma_t \), and \( \kappa \) such that \( Y \) satisfies (20.25). Using the formula for the constant \( c_n \) derived in the previous exercise, specify a condition on the parameters \( A_t, \alpha, \mu_t, \) and \( \sigma_t \) that is equivalent to \( \delta > 0 \).

**20.3.** Use the dynamics (20.25) of \( Y \) and the definition (20.30) of \( \chi \) to verify the formula (20.35) for the risk of a stock return in the model of perfect competition in Section 20.5. (p.550)

Notes:

1. More precisely, marginal \( q \) is the marginal value of capital divided by the price of capital, but we take capital to be the numeraire, so its price is 1. Likewise, average \( q \) is the value of the firm \( J(K, X) \) divided by the replacement cost of capital. The replacement cost of capital equals \( K_n \), again because we take capital as the numeraire.

2. An equivalent model is sometimes studied. In that model, investment is measured by its cost rather than by its contribution to capital. So, using a \( \hat{\cdot} \) to distinguish the models, the cost of investment is called \( \hat{I} \) rather than \( \theta(K, I) \), and the capital evolution equation is \( K_t = f(K_t, \hat{I}_t) - \rho K_t \) for some function \( f \) that should be assumed to be strictly increasing in \( \hat{I} \). Given such a model, we can define \( I_t = f(K_t, \hat{I}_t) \) and \( \theta(K, I) = \hat{I}_t = f(K_t, \cdot)^{-1}(I_t) \) to write the model in the form studied here.

3. Condition (20.13) is violated when the standard deviation \( \sigma_t \) of \( \chi \) is large or the risk-neutral expected growth rate \( r - \delta \) of \( \chi \) is large or the interest rate \( r \) is small or when \( \eta \) is large (close to 1). It is quite intuitive that high risk or high expected growth of \( \chi \) or a low interest rate produces a high value for growth options. To see how \( \eta \) is related to the value of growth options, recall that, in the monopoly model with Cobb-Douglas production, the parameter \( \eta \) depends on returns to scale and on the elasticity of demand. A high value for \( \eta \) means that returns to scale diminish slowly and/or demand is close to
being perfectly elastic. Either condition implies that marginal
operating cash flows decrease slowly as capital increases.

(4.) To see this, set \( s = \frac{x}{k^{1-\eta}} \) and \( s' = \beta (1 - \beta) \). Then,

\[
J_d(k, x) = s - (s \frac{s'^2}{\beta - 1}) (1 - s)^{1-\eta}
\]

At the boundary, we have \( s = s' \) and \( J_d(k, x) = s^\eta (1 - s)^{1-\eta} = 1 \). Below the
boundary, \( s < s' \), and

\[
\frac{\partial}{\partial s} \left[ s - (s \frac{s'^2}{\beta - 1}) (1 - s)^{1-\eta} \right] = 1 - (s \frac{s'^2}{\beta - 1}) (1 - s)^{\eta} > 0.
\]

Therefore, \( J_d(k, x) \) is increasing in \( x \) below the boundary, increasing
up to its value of 1 at the boundary.

(5.) The real meaning of the condition \( (J_1 - 1) dK = 0 \) is that, with
probability 1,

\[
(\forall \theta) \quad \int_0^x J_d(K, x, \theta) - 1 dK = 0.
\]

(6.) If there were decreasing returns to scale, then free entry
would imply that there should be infinitely many firms of
infinitely small scale, which is certainly unrealistic. It may be
more realistic to assume increasing returns to scale up to
some point and then decreasing returns thereafter, yielding a
U-shaped average cost curve, but constant returns to scale is
more tractable.

(7.) By (20.18) and iterated expectations,

\[
\mathbb{E} \left[ \frac{M_u}{M_t} C_{su} \right] = \mathbb{E} \left[ \frac{M_{u-1}}{M_{t-1}} \mathbb{E}_{\theta} \left[ \frac{M_u}{M_{t+1}} C_{su} \right] \right] = e^{\theta_1} \mathbb{E} \left[ \frac{M_{u-1}}{M_{t-1}} \mathbb{E}_{\theta} \left[ \frac{M_u}{M_{t+1}} e^{\theta_1} \right] \right]
\]

\[
= e^{\theta_1} \mathbb{E} \left[ \frac{M_{u-1}}{M_{t-1}} \mathbb{E}_{\theta} \left[ \frac{M_u}{M_{t+1}} \right] \right] = e^{\theta_1} \mathbb{E} \left[ \frac{M_u}{M_t} \right].
\]