Part V

Additional Topics in Asset Pricing

Chapter 15

Behavioral Finance and Asset Pricing

This chapter considers asset pricing when investors' asset demands incorporate some elements of irrationality. Irrationality can occur because investors' preferences are subject to psychological biases or because investors make systematic errors in judging the probability distribution of asset returns. Incorporating irrationality is a departure from von Neumann-Morgenstern expected utility maximization and the standard or classical economic approach. A model that incorporates some form of irrationality is unlikely to be useful for drawing normative conclusions regarding an individual's asset choice. Rather, such a model attempts to provide a positive or descriptive theory of how individuals actually behave. For this reason, the approach is referred to as "behavioral finance."

There is both experimental evidence as well as conventional empirical research documenting investor behavior that is inconsistent with von Neumann-Morgenstern expected utility theory. Numerous forms of cognitive biases and judgement errors appear to characterize the preferences of at least some individuals. Surveys by (Hirshleifer 2001), (Daniel, Hirshleifer, and Teoh 2001), and (Barberis and Thaler 2002) describe the evidence for these behavioral phenomena. However, to date there have been relatively few models that analyze how irrationality might affect equilibrium asset prices. This chapter examines two recent behavioral asset pricing models.

The first is an intertemporal consumption and portfolio choice model by Nicholas Barberis, Ming Huang, and Jesus Santos (Barberis, Huang, and Santos 2001) that incorporates two types of biases that are prominent in the behavioral finance literature. They are *loss aversion* and the *house money effect*. These biases fall within the general category of *prospect theory*. Prospect theory deviates from von Neumann-Morgenstern expected utility maximization because investor utility is a function of recent changes in, rather than simply the current level of, financial wealth. In particular, investor utility characterized by prospect theory may be more sensitive to recent losses than recent gains in financial wealth, this phenomenon being referred to as *loss aversion*. Moreover, losses following previous losses create more disutility than losses following previous gains. After a run-up in asset prices, the investor is less risk averse because subsequent losses would be "cushioned" by the previous gains. This is the so-called *house money* effect.[1]

An implication of this intertemporal variation in risk aversion is that after a substantial rise in asset prices, lower investor risk aversion can drive prices even higher. Hence, asset prices display volatility that is greater than that predicted by observed changes in fundamentals, such as changes in dividends. This also generates predictability in asset returns. A substantial recent fall (*rise*) in asset prices increases (*decreases*) risk aversion and expected asset returns. It can also

[1] This expression derives from the psychological misperception that a gambler's (unexpected) winnings are the casino house's money. The gambler views these winnings as different from his initial wealth upon entering the casino. Hence, the gambler is willing to bet more aggressively in the future because if the house's money is lost, the disutility of this loss will be small relative to the disutility of losing the same amount of his initial wealth.

imply a high equity risk premium because the "excess" volatility in stock prices leads loss-averse investors to demand a relatively high average rate of return on stocks.

Prospect theory assumes that investors are overly concerned with changes in financial wealth measured against some reference points, such as profits or losses measured from the times when assets were first purchased. They care about these holding period gains or losses more than would be justified by their effects on consumption, and this influences their risk-taking behavior. This psychological concept was advanced by Daniel Kahneman and Amos Tversky (Kahneman and Tversky 1979) and is based primarily on experimental evidence.[2] For example, Richard Thaler and Eric Johnson (Thaler and Johnson 1990) find that individuals faced with a sequence of gambles are more willing to take risk if they have made gains from previous gambles, evidence consistent with the house money effect. However, in a recent study of the behavior of traders of the Chicago Board of Trade's Treasury bond futures, Joshua Coval and Tyler Shumway (Coval and Shumway 2003) find evidence consistent with loss aversion but not the house money effect.

The second model presented in this chapter examines how equilibrium asset prices are affected when some investors are rational but others suffer from systematic optimism or pessimism. Leonid Kogan, Stephen Ross, Jiang Wang, and Mark Westerfied (Kogan, Ross, Wang, and Westerfield 2006) construct a simple endowment economy where rational and irrational investors are identical except that the irrational investors systematically misperceive the expected growth rate of the aggregate dividend process. Interestingly, it is shown that this economy can be transformed into one where the irrational traders can be viewed as acting rationally but their utilities are state dependent. This transformation of the problem allows it to be solved using standard techniques.

[2] Daniel Kahneman was awarded the Nobel prize in economics in 2002.

Kogan, Ross, Wang, and Westerfield's general equilibrium model shows that investors having irrational beliefs regarding the economy's fundamentals may not necessarily lose wealth to rational investors and be driven out of the asset market.[3] Moreover, in those instances where irrational individuals do lose wealth relative to the rational individuals, so that they do not survive in the long run, their trading behavior can significantly affect asset prices for substantial periods of time.

We now turn to the Barberis, Huang, and Santos model, which generalizes a standard consumption and portfolio choice problem to incorporate aspects of prospect theory.

15.1 The Effects of Psychological Biases on Asset Prices

The Barberis, Huang, and Santos model is based on the following assumptions.

15.1.1 Assumptions

In the discussion that follows, the model economy has the following characteristics.

Technology

A discrete-time endowment economy is assumed. The risky asset (or a portfolio of all risky assets) pays a stream of dividends in the form of perishable output. Denote the date t amount of this dividend as D_t. In the Economy I version of the Barberis, Huang, and Santos model, it is assumed that aggregate consumption equals dividends. This is the standard Lucas economy assumption

[3] This result was shown by Bradford De Long, Andrei Shleifer, Lawrence Summers, and Robert Waldmann (DeLong, Shleifer, Summers, and Waldmann 1991) in a partial equilibrium model.

(Lucas 1978). However, in the Economy II version of their model, which will be the focus of our analysis, the risky asset's dividends are distinct from aggregate consumption due to the assumed existence of nonfinancial, or labor, income.[4] Recall that we studied this labor income extension of the standard Lucas economy in Chapter 6. Nonfinancial wealth can be interpreted as human capital and its dividend as labor income. Thus, in equilibrium, aggregate consumption, $\overline{C}_t$, equals dividends, D_t, plus nonfinancial income, Y_t, because both dividends and nonfinancial income are assumed to be perishable. Aggregate consumption and dividends are assumed to follow the joint lognormal process

$$\begin{aligned} \ln\left(\overline{C}_{t+1}/\overline{C}_t\right) &= g_C + \sigma_C \eta_{t+1} \\ \ln\left(D_{t+1}/D_t\right) &= g_D + \sigma_D \varepsilon_{t+1} \end{aligned} \tag{15.1}$$

where the error terms are serially uncorrelated and distributed as

$$\begin{pmatrix} \eta_t \\ \varepsilon_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right) \tag{15.2}$$

The return on the risky asset from date t to date $t+1$ is denoted R_{t+1}. A one-period risk-free investment is assumed to be in zero net supply, and its return from date t to date $t+1$ is denoted $R_{f,t}$.[5] The equilibrium value for $R_{f,t}$ is derived next.

Preferences

[4]Note that in a standard endowment economy, consumption and dividends are perfectly correlated since they equal each other in equilibrium. Empirically, it is obvious that aggregate consumption does not equal, nor is perfectly correlated with, aggregate stock dividends. Hence, to make the model more empirically relevant, the Economy II version of the model introduces nonfinancial income, which avoids the implication of perfect correlation.

[5]Since the risk-free asset is in zero net supply, the representative individual's equilibrium holding of this asset is zero. Similar to the case of the Cox, Ingersoll, and Ross model presented in Chapter 13, $R_{f,t}$ is interpreted as the shadow riskless return.

Representative, infinitely lived individuals maximize lifetime utility of the form

$$E_0\left[\sum_{t=0}^{\infty}\left(\delta^t\frac{C_t^\gamma}{\gamma}+b_t\delta^{t+1}v\left(X_{t+1},w_t,z_t\right)\right)\right] \tag{15.3}$$

where C_t is the individual's consumption at date t, $\gamma < 1$, and δ is a time discount factor. w_t denotes the number of shares of the risky asset held by the individual at date t. X_{t+1} is defined as the total excess return or gain that the individual earned from holding the risky asset between date t and date $t+1$. Specifically, this risky-asset gain is assumed to be measured relative to the alternative of holding wealth in the risk-free asset and is given by

$$X_{t+1} \equiv w_t\left(R_{t+1}-R_{f,t}\right) \tag{15.4}$$

z_t is a measure of the individual's prior gains as a fraction of w_t. $z_t < (>)\, 1$ denotes a situation in which the investor has earned prior gains (*losses*) on the risky asset. The prior gain factor, z_t, is assumed to follow the process

$$z_t = (1-\eta)+\eta z_{t-1}\frac{\overline{R}}{R_t} \tag{15.5}$$

where $0 \leq \eta \leq 1$ and $\overline{R}$ is a parameter, approximately equal to the average risky-asset return, that makes the steady state value of z_t equal 1. If $\eta = 0$, z_t $= 1$ for all t. At the other extreme, when $\eta = 1$, z_t is smaller than $z_{t-1}$when risky-asset returns were relatively high last period, $R_t > \overline{R}$. Conversely, when $\eta = 1$ but $R_t < \overline{R}$, z_t is larger than z_{t-1}. For intermediate cases of $0 < \eta < 1$, z_t adjusts partially to prior asset returns. In general, the greater η is, the longer the investor's memory in measuring prior gains from the risky asset.

The function $v\left(\cdot\right)$ characterizes the prospect theory effect of risky-asset gains on utility.[6] For the case of $z_t = 1$ (no prior gains or losses), this function displays pure loss aversion:

$$v\left(X_{t+1}, w_t, 1\right) = \begin{cases} X_{t+1} & \text{if } X_{t+1} \geq 0 \\ \lambda X_{t+1} & \text{if } X_{t+1} < 0 \end{cases} \tag{15.6}$$

where $\lambda > 1$. Hence, ceteris paribus, losses have a disproportionately bigger impact on utility. When $z_t \neq 1$, the function $v(\cdot)$ reflects prospect theory's house money effect. In the case of prior gains ($z_t \leq 1$), the function takes the form

$$v\left(X_{t+1}, w_t, z_t\right) = \begin{cases} X_{t+1} & \text{if } R_{t+1} \geq z_t R_{f,t} \\ X_{t+1} + (\lambda - 1) w_t \left(R_{t+1} - z_t R_{f,t}\right) & \text{if } R_{t+1} < z_t R_{f,t} \end{cases} \tag{15.7}$$

The interpretation of this function is that when a return exceeds the cushion built by prior gains, that is, $R_{t+1} \geq z_t R_{f,t}$, it affects utility one-for-one. However, when the gain is less than the amount of prior gains, $R_{t+1} < z_t R_{f,t}$, it has a greater than one-for-one impact on disutility. In the case of prior losses ($z_t > 1$), the function becomes

$$v\left(X_{t+1}, w_t, z_t\right) = \begin{cases} X_{t+1} & \text{if } X_{t+1} \geq 0 \\ \lambda\left(z_t\right) X_{t+1} & \text{if } X_{t+1} < 0 \end{cases} \tag{15.8}$$

where $\lambda\left(z_t\right) = \lambda + k\left(z_t - 1\right)$, $k > 0$. Here we see that losses that follow previous losses are penalized at the rate of $\lambda\left(z_t\right)$, which exceeds λ and grows larger as prior losses become larger (z_t exceeds unity).

Finally, the prospect theory term in the utility function is scaled to make the risky-asset price-dividend ratio and the risky-asset risk premium stationary

[6] Since $v(\cdot)$ depends only on the risky asset's returns, it is assumed that the individual is not subject to loss aversion on nonfinancial assets.

variables as aggregate wealth increases over time.[7] The form of this scaling factor is chosen to be

$$b_t = b_0 \overline{C}_t^{\gamma - 1} \tag{15.9}$$

where $b_0 > 0$ and $\overline{C}_t$ is aggregate consumption at date t.[8]

15.1.2 Solving the Model

The state variables for the individual's consumption-portfolio choice problem are wealth, W_t, and z_t. Intuitively, since the aggregate consumption - dividend growth process in equation (15.1) is an independent, identical distribution, the dividend level is not a state variable. We start by assuming that the ratio of the risky-asset price to its dividend is a function of only the state variable z_t; that is, $f_t \equiv P_t/D_t = f_t(z_t)$, and then show that an equilibrium exists in which this is true.[9] Given this assumption, the return on the risky asset can be written as

$$\begin{aligned} R_{t+1} &= \frac{P_{t+1} + D_{t+1}}{P_t} = \frac{1 + f(z_{t+1})}{f(z_t)} \frac{D_{t+1}}{D_t} \\ &= \frac{1 + f(z_{t+1})}{f(z_t)} e^{g_D + \sigma_D \varepsilon_{t+1}} \end{aligned} \tag{15.10}$$

It is also assumed that an equilibrium exists in which the risk-free return is constant; that is, $R_{f,t} = R_f$. This will be verified by the solution to the agent's first-order conditions. Making this assumption simplifies the form of the function v. From (15.7) and (15.8) it can be verified that v is proportional to w_t.

---

[7] Without the scaling factor, as wealth (output) grows at rate g_D, the prospect theory term would dominate the conventional constant relative-risk-aversion term.

[8] Because $\overline{C}_t$ is assumed to be aggregate consumption, the individual views b_t as an exogeneous variable.

[9] This is plausible because the standard part of the utility function displays constant relative risk aversion. With this type of utility, optimal portfolio proportions would not be a function of wealth.

Hence, $v(X_{t+1}, w_t, z_t)$ can be written as $v(X_{t+1}, w_t, z_t) = w_t \widehat{v}(R_{t+1}, z_t)$, where for $z_t < 1$

$$\widehat{v}(R_{t+1}, z_t) = \begin{cases} R_{t+1} - R_f & \text{if } R_{t+1} \geq z_t R_f \\ R_{t+1} - R_f + (\lambda - 1)(R_{t+1} - z_t R_f) & \text{if } R_{t+1} < z_t R_f \end{cases} \tag{15.11}$$

and for $z_t > 1$

$$\widehat{v}(R_{t+1}, z_t) = \begin{cases} R_{t+1} - R_f & \text{if } R_{t+1} \geq R_f \\ \lambda(z_t)(R_{t+1} - R_f) & \text{if } R_{t+1} < R_f \end{cases} \tag{15.12}$$

The individual's maximization problem is then

$$\max_{\{C_t, w_t\}} E_0 \left[\sum_{t=0}^{\infty} \left(\delta^t \frac{C_t^\gamma}{\gamma} + b_0 \delta^{t+1} \overline{C}_t^{\gamma-1} w_t \widehat{v}(R_{t+1}, z_t) \right) \right] \tag{15.13}$$

subject to the budget constraint

$$W_{t+1} = (W_t + Y_t - C_t) R_f + w_t (R_{t+1} - R_f) \tag{15.14}$$

and the dynamics for z_t given in (15.5). Define $\delta^t J(W_t, z_t)$ as the derived utility-of-wealth function. Then the Bellman equation for this problem is

$$J(W_t, z_t) = \max_{\{C_t, w_t\}} \frac{C_t^\gamma}{\gamma} + E_t \left[b_0 \delta \overline{C}_t^{\gamma-1} w_t \widehat{v}(R_{t+1}, z_t) + \delta J(W_{t+1}, z_{t+1}) \right] \tag{15.15}$$

Taking the first-order conditions with respect to C_t and w_t, one obtains

$$0 = C_t^{\gamma-1} - \delta R_f E_t \left[J_W(W_{t+1}, z_{t+1}) \right] \tag{15.16}$$

$$\begin{aligned} 0 &= E_t\left[b_0\overline{C}_t^{\gamma-1}\widehat{v}\left(R_{t+1},z_t\right)+J_W\left(W_{t+1},z_{t+1}\right)\left(R_{t+1}-R_f\right)\right] \\ &= b_0\overline{C}_t^{\gamma-1}E_t\left[\widehat{v}\left(R_{t+1},z_t\right)\right]+E_t\left[J_W\left(W_{t+1},z_{t+1}\right)R_{t+1}\right] \\ &\quad -R_f E_t\left[J_W\left(W_{t+1},z_{t+1}\right)\right] \end{aligned} \tag{15.17}$$

It is straightforward (and left as an end-of-chapter exercise) to show that (15.16) and (15.17) imply the standard envelope condition

$$C_t^{\gamma-1} = J_W\left(W_t, z_t\right) \tag{15.18}$$

Substituting this into (15.16), one obtains the Euler equation

$$1 = \delta R_f E_t\left[\left(\frac{C_{t+1}}{C_t}\right)^{\gamma-1}\right] \tag{15.19}$$

Using (15.18) and (15.19) in (15.17) implies

$$\begin{aligned} 0 &= b_0\overline{C}_t^{\gamma-1}E_t\left[\widehat{v}\left(R_{t+1},z_t\right)\right]+E_t\left[C_{t+1}^{\gamma-1}R_{t+1}\right]-R_f E_t\left[C_{t+1}^{\gamma-1}\right] \\ &= b_0\overline{C}_t^{\gamma-1}E_t\left[\widehat{v}\left(R_{t+1},z_t\right)\right]+E_t\left[C_{t+1}^{\gamma-1}R_{t+1}\right]-C_t^{\gamma-1}/\delta \end{aligned} \tag{15.20}$$

or

$$1 = b_0\left(\frac{\overline{C}_t}{C_t}\right)^{\gamma-1}\delta E_t\left[\widehat{v}\left(R_{t+1},z_t\right)\right]+\delta E_t\left[R_{t+1}\left(\frac{C_{t+1}}{C_t}\right)^{\gamma-1}\right] \tag{15.21}$$

In equilibrium, conditions (15.19) and (15.21) hold with the representative agent's consumption, C_t, replaced with aggregate consumption, $\overline{C}_t$. Using the assumption in (15.1) that aggregate consumption is lognormally distributed, we can compute the expectation in (15.19) to solve for the risk-free interest rate:

$$R_f = e^{(1-\gamma)g_C-\frac{1}{2}(1-\gamma)^2\sigma_C^2}/\delta \tag{15.22}$$

Using (15.1) and (15.10), condition (15.21) can also be simplified:

$$1 = b_0 \delta E_t \left[\widehat{v}\left(R_{t+1}, z_t\right) \right] + \delta E_t \left[\frac{1 + f\left(z_{t+1}\right)}{f\left(z_t\right)} e^{g_D + \sigma_D \varepsilon_{t+1}} \left(e^{g_C + \sigma_C \eta_{t+1}}\right)^{\gamma - 1} \right] \tag{15.23}$$

or

$$\begin{aligned} 1 \quad = \quad & b_0 \delta E_t \left[\widehat{v}\left(\frac{1 + f\left(z_{t+1}\right)}{f\left(z_t\right)} e^{g_D + \sigma_D \varepsilon_{t+1}}, z_t \right) \right] \\ & + \delta e^{g_D - (1-\gamma) g_C + \frac{1}{2}(1-\gamma)^2 \sigma_C^2 \left(1 - \rho^2\right)} E_t \left[\frac{1 + f\left(z_{t+1}\right)}{f\left(z_t\right)} e^{\left(\sigma_D - (1-\gamma)\rho \sigma_C\right) \varepsilon_{t+1}} \right] \end{aligned} \tag{15.24}$$

The price-dividend ratio, $P_t / D_t = f_t\left(z_t\right)$, can be computed numerically from (15.24). However, because $z_{t+1} = 1 + \eta\left(z_t \frac{\overline{R}}{R_{t+1}} - 1\right)$ and $R_{t+1} = \frac{1 + f(z_{t+1})}{f(z_t)} e^{g_D + \sigma_D \varepsilon_{t+1}}$, z_{t+1} depends upon z_t, $f\left(z_t\right)$, $f\left(z_{t+1}\right)$, and ε_{t+1}; that is,

$$z_{t+1} = 1 + \eta\left(z_t \frac{\overline{R} f\left(z_t\right) e^{-g_D - \sigma_D \varepsilon_{t+1}}}{1 + f\left(z_{t+1}\right)} - 1 \right) \tag{15.25}$$

Therefore, (15.24) and (15.25) need to be solved jointly. Barberis, Huang, and Santos describe an iterative numerical technique for finding the function $f(\cdot)$. Given all other parameters, they guess an initial function, $f^{(0)}$, and then use it to solve for z_{t+1} in (15.25) for given z_t and ε_{t+1}. Then, they find a new candidate solution, $f^{(1)}$, using the following recursion that is based on (15.24):

$$\begin{aligned} f^{(i+1)}\left(z_t\right) \quad = \quad & \delta e^{g_D - (1-\gamma) g_C + \frac{1}{2}(1-\gamma)^2 \sigma_C^2 \left(1 - \rho^2\right)} \times \\ & E_t \left[\left[1 + f^{(i)}\left(z_{t+1}\right) \right] e^{\left(\sigma_D - (1-\gamma)\rho \sigma_C\right) \varepsilon_{t+1}} \right] \\ & + f^{(i)}\left(z_t\right) b_0 \delta E_t \left[\widehat{v}\left(\frac{1 + f^{(i)}\left(z_{t+1}\right)}{f^{(i)}\left(z_t\right)} e^{g_D + \sigma_D \varepsilon_{t+1}}, z_t \right) \right], \ \forall z_t \end{aligned} \tag{15.26}$$

where the expectations are computed using a Monte Carlo simulation of the ε_{t+1}. Given the new candidate function, $f^{(1)}$, z_{t+1} is again found from (15.25). The procedure is repeated until the function $f^{(i)}$ converges.

15.1.3 Model Results

For reasonable parameter values, Barberis, Huang, and Santos find that $P_t/D_t = f_t(z_t)$ is a decreasing function of z_t. The intuition was described earlier: if there were prior gains from holding the risky asset (z_t is low), then investors become less risk averse and bid up the price of the risky asset.

Using their estimate of $f(\cdot)$, the unconditional distribution of stock returns is simulated from a randomly generated sequence of ε_t's. Because dividends and consumption follow separate processes and stock prices have volatility exceeding that of dividend fundamentals, the volatility of stock prices can be made substantially higher than that of consumption. Moreover, because of loss aversion, the model can generate a significant equity risk premium for reasonable values of the consumption risk aversion parameter γ. Thus, the model provides an explanation for the "equity premium puzzle." Because the investor cares about stock volatility, per se, a large premium can exist even though stocks may not have a high correlation with consumption.[10]

The model also generates predictability in stock returns: returns tend to be higher following crashes (when z_t is high) and smaller following expansions (when z_t is low). An implication of this is that stock returns are negatively correlated at long horizons, a feature documented by empirical research such as (Fama and French 1988), (Poterba and Summers 1988), and (Richards 1997).

The Barberis, Huang, and Santos model is one with a single type of representative individual who suffers from psychological biases. The next model that

[10] Recall that in standard consumption asset pricing models, an asset's risk premium depends only on its return's covariance with consumption.

we consider assumes that there are two types of representative individuals, those with rational beliefs and those with irrational beliefs regarding the economy's fundamentals. Important insights are obtained by analyzing the interactions of these two groups of investors.

15.2 The Impact of Irrational Traders on Asset Prices

The Kogan, Ross, Wang, and Westerfield model is based on the following assumptions.

15.2.1 Assumptions

The model is a simplified endowment economy with two different types of representative individuals, where one type suffers from either irrational optimism or pessimism regarding risky-asset returns. Both types of individuals maximize utility of consumption at a single, future date.[11]

Technology

There is a risky asset that represents a claim on a single, risky dividend payment made at the future date $T > 0$. The value of this dividend payment is denoted D_T, and it is the date T realization of the geometric Brownian motion process

$$dD_t/D_t = \mu dt + \sigma dz \tag{15.27}$$

where μ and σ are constants, $\sigma > 0$, and $D_0 = 1$. Note that while the process in equation (15.27) is observed at each date $t \in [0, T]$, only its realization at date T determines the risky asset's single dividend payment, D_T. As with

[11] Alvaro Sandroni (Sandroni 2000) developed a discrete-time model with similar features that allows the different types of individuals to consume at multiple future dates.

other endowment economies, it is assumed that the date T dividend payment is perishable output so that, in equilibrium, it equals aggregate consumption, $C_T = D_T$.

Also, it is assumed that there is a market for risk-free borrowing or lending where payment occurs with certainty at date T. In other words, individuals can buy or sell (issue) a zero-coupon bond that makes a default-free payment of 1 at date T. This bond is assumed to be in zero net supply; that is, the aggregate net amount of risk-free lending or borrowing is zero. However, because there are heterogeneous groups of individuals in the economy, some individuals may borrow while others will lend.

Preferences

All individuals in the economy have identical constant relative-risk-aversion utility defined over their consumption at date T. However, there are two different groups of representative individuals. The first group of individuals are rational traders who have a date 0 endowment equal to one-half of the risky asset and maximize the expected utility function

$$E_0\left[\frac{C_{r,T}^{\gamma}}{\gamma}\right] \tag{15.28}$$

where $C_{r,T}$ is the date T consumption of the rational traders and $\gamma < 1$. The second group of individuals are irrational traders. They also possess a date 0 endowment of one-half of the risky asset but incorrectly believe that the probability measure is different from the actual one. Rather than thinking that the aggregate dividend process is given by (15.27), the irrational traders incorrectly perceive the dividend process to be

$$dD_t/D_t = \left(\mu + \sigma^2\eta\right)dt + \sigma d\widehat{z} \tag{15.29}$$

where the irrational traders believe $d\widehat{z}$ is a Brownian motion, whereas in reality, $d\widehat{z} = dz - \sigma\eta dt$. The irrationality parameter, η, is assumed to be a constant. A positive value of η implies that the irrational individuals are too optimistic about the risky asset's future dividend payment, while a negative value of η indicates pessimism regarding the risky asset's payoff. Hence, rather than believe that the probability measure P is generated by the Brownian motion process dz, irrational traders believe that the probability measure is generated by $d\widehat{z}$, which we refer to as the probability measure $\widehat{P}$.[12] Therefore, an irrational individual's expected utility is

$$\widehat{E}_0\left[\frac{C_{n,T}^{\gamma}}{\gamma}\right] \tag{15.30}$$

where $C_{n,T}$ is the date T consumption of the irrational trader.

15.2.2 Solution Technique

We start by showing that the irrational individual's utility can be reinterpreted as the state-dependent utility of a rational individual. Recall from Chapter 10 that as a result of Girsanov's theorem, a transformation of the type $d\widehat{z} = dz - \sigma\eta dt$ leads to $\widehat{P}$ and P being equivalent probability measures and that there exists a sequence of strictly positive random variables, ξ_t, that can transform one distribution to the other. Specifically, recall from equation (10.11) that Girsanov's theorem implies $d\widehat{P}_T = (\xi_T/\xi_0)\, dP_T$, where based on (10.12)

$$\begin{aligned} \xi_T &= \exp\left[\int_0^T \sigma\eta dz - \frac{1}{2}\int_0^T (\sigma\eta)^2\, ds\right] \\ &= e^{-\frac{1}{2}\sigma^2\eta^2 T + \sigma\eta(z_T - z_0)} \end{aligned} \tag{15.31}$$

[12] It should be emphasized that the probability measure $\widehat{P}$ is not necessarily the risk-neutral probability measure. The dividend process is not an asset return process so that μ is not an asset's expected rate of return and η is not a risk premium.

and where, without loss of generality, we have assumed that $\xi_0 = 1$. The second line in (15.31) follows because σ and η are assumed to be constants, implying that ξ_t follows the lognormal process $d\xi/\xi = \sigma\eta dz$. Similar to (10.30), an implication of $d\widehat{P}_T = \xi_T dP_T$ is that an irrational trader's expected utility can be written as

$$\begin{aligned} \widehat{E}_0\left[\frac{C_{n,T}^{\gamma}}{\gamma}\right] &= E_0\left[\xi_T \frac{C_{n,T}^{\gamma}}{\gamma}\right] \\ &= E_0\left[e^{-\frac{1}{2}\sigma^2\eta^2 T + \sigma\eta(z_T - z_0)} \frac{C_{n,T}^{\gamma}}{\gamma}\right] \end{aligned} \tag{15.32}$$

From (15.32) we see that the objective function of the irrational trader is observationally equivalent to that of a rational trader whose utility is state dependent. The state variable affecting utility, the Brownian motion z_T, is the same source of uncertainty determining the risky asset's dividend payment.

While the ability to transform the behavior of an irrational individual to that of a rational one may depend on the particular way that irrationality is modeled, this transformation allows us to use standard methods for determining the economy's equilibrium. Given the assumption of two different groups of representative individuals, we can solve for an equilibrium where the representative individuals act competitively, taking the price of the risky asset and the risk-free borrowing or lending rate as given. In addition, because there is only a single source of uncertainty, that being the risky asset's payoff, the economy is dynamically complete.

Given market completeness, let us apply the martingale pricing method introduced in Chapter 12. Each individual's lifetime utility function can be interpreted as of the form of (12.55) but with interim utility of consumption equaling zero and only a utility of terminal bequest being nonzero. Hence, based on equation (12.57), the result of each individual's static optimization is that his terminal marginal utility of consumption is proportional to the pricing kernel:

$$C_{r,T}^{\gamma-1} = \lambda_r M_T \tag{15.33}$$

$$\xi_T C_{n,T}^{\gamma-1} = \lambda_n M_T \tag{15.34}$$

where λ_r and λ_n are the Lagrange multipliers for the rational and irrational individuals, respectively. Substituting out for M_T, we can write

$$C_{r,T} = (\lambda \xi_T)^{-\frac{1}{1-\gamma}} C_{n,T} \tag{15.35}$$

where we define $\lambda \equiv \lambda_r / \lambda_n$. Also note that the individuals' terminal consumption must sum to the risky asset's dividend payment

$$C_{r,T} + C_{n,T} = D_T \tag{15.36}$$

Equations (15.35) and (15.36) allow us to write each individual's terminal consumption as

$$C_{r,T} = \frac{1}{1 + (\lambda \xi_T)^{\frac{1}{1-\gamma}}} D_T \tag{15.37}$$

Substituting (15.37) into (15.35), we also obtain

$$C_{n,T} = \frac{(\lambda \xi_T)^{\frac{1}{1-\gamma}}}{1 + (\lambda \xi_T)^{\frac{1}{1-\gamma}}} D_T \tag{15.38}$$

Similar to what was done in Chapter 13, the parameter $\lambda = \lambda_r / \lambda_n$ is determined by the individuals' initial endowments of wealth. Each individual's initial wealth is an asset that pays a dividend equal to the individual's terminal consumption. To value this wealth, we must determine the form of the stochastic discount factor used to discount consumption. As a prelude, note that the date t price of the zero coupon bond that pays 1 at date $T > t$ is given by

$$P(t,T) = E_t\left[M_T/M_t\right] \tag{15.39}$$

In what follows, we deflate all asset prices, including the individuals' initial wealths, by this zero-coupon bond price. This is done for analytical convenience, though it should be noted that using the zero-coupon bond as the numeraire is somewhat different from using the value of a money market investment as the numeraire, as was done in Chapter 10. While the return on the zero-coupon bond over its remaining time to maturity is risk-free, its instantaneous return will not, in general, be risk-free.

Let us define $W_{r,0}$ and $W_{n,0}$ as the initial wealths, deflated by the zero-coupon bond price, of the rational and irrational individuals, respectively. They equal

$$\begin{aligned}
W_{r,0} &= \frac{E_0\left[C_{r,T}M_T/M_0\right]}{E_0\left[M_T/M_0\right]} = \frac{E_0\left[C_{r,T}M_T\right]}{E_0\left[M_T\right]} \\
&= \frac{E_0\left[C_{r,T}C_{r,T}^{\gamma-1}/\lambda_r\right]}{E_0\left[C_{r,T}^{\gamma-1}/\lambda_r\right]} = \frac{E_0\left[C_{r,T}^{\gamma}\right]}{E_0\left[C_{r,T}^{\gamma-1}\right]} \\
&= \frac{E_0\left[\left[1+(\lambda\xi_T)^{\frac{1}{1-\gamma}}\right]^{-\gamma}D_T^{\gamma}\right]}{E_0\left[\left[1+(\lambda\xi_T)^{\frac{1}{1-\gamma}}\right]^{1-\gamma}D_T^{\gamma-1}\right]}
\end{aligned} \tag{15.40}$$

where in the second line of (15.40) we used (15.33) to substitute for M_T and then in the third line we used (15.37) to substitute for $C_{r,T}$. A similar derivation that uses (15.34) and (15.38) leads to

$$W_{n,0} = \frac{E_0\left[(\lambda\xi_T)^{\frac{1}{1-\gamma}}\left[1+(\lambda\xi_T)^{\frac{1}{1-\gamma}}\right]^{-\gamma}D_T^{\gamma}\right]}{E_0\left[\left[1+(\lambda\xi_T)^{\frac{1}{1-\gamma}}\right]^{1-\gamma}D_T^{\gamma-1}\right]} \tag{15.41}$$

Because it was assumed that the rational and irrational individuals are each initially endowed with equal one-half shares of the risky asset, then it must be the case that $W_{r,0} = W_{n,0}$. Equating the right-hand sides of equations (15.40) and (15.41) determines the value for λ. The expectations in these equations can be computed by noting that ξ_T satisfies (15.31) and is lognormally distributed and that

$$D_T/D_t = e^{\left[\mu - \frac{1}{2}\sigma^2\right](T-t)+\sigma(z_T - z_t)} \tag{15.42}$$

and is also lognormally distributed.[13] It is left as an end-of-chapter exercise to verify that the value of λ that solves the equality $W_{r,0} = W_{n,0}$ is given by

$$\lambda = e^{-\gamma\eta\sigma^2 T} \tag{15.43}$$

Given this value of λ, we have now determined the form of the pricing kernel and can solve for the equilibrium price of the risky asset. Define S_t as the date $t < T$ price of the risky asset deflated by the price of the zero-coupon bond. Then if we also define $\varepsilon_{T,t} \equiv \lambda\xi_T = \xi_t e^{-\gamma\eta\sigma^2 T - \frac{1}{2}\sigma^2\eta^2(T-t)+\sigma\eta(z_T - z_t)}$, the deflated risky-asset price can be written as

$$S_t = \frac{E_t\left[D_T M_T/M_t\right]}{E_t\left[M_T/M_t\right]} = \frac{E_t\left[\left(1+\varepsilon_{T,t}^{\frac{1}{1-\gamma}}\right)^{1-\gamma} D_T^{\gamma}\right]}{E_t\left[\left(1+\varepsilon_{T,t}^{\frac{1}{1-\gamma}}\right)^{1-\gamma} D_T^{\gamma-1}\right]} \tag{15.44}$$

While it is not possible to characterize in closed form the rational and irrational individuals' portfolio policies, we can still derive insights regarding equilibrium asset pricing.[14]

[13] Recall that it was assumed that $D_0 = 1$. Note also that powers of ξ_T and D_T, such as D_T^{γ}, are also lognormally distributed.

[14] Kogan, Ross, Wang, and Westerfield show that the individuals' demand for the risky asset, ω, satisfies the bound $|\omega| \leq 1 + |\eta|\left(2-\gamma\right)/\left(1-\gamma\right)$.

15.2.3 Analysis of the Results

For the limiting case of there being only rational individuals, that is, $\eta = 0$, then $\varepsilon_{T,t} = \xi_t = 1$ and from (15.44) the deflated stock price, $S_{r,t}$, is

$$\begin{aligned} S_{r,t} &= \frac{E_t\left[D_T^{\gamma}\right]}{E_t\left[D_T^{\gamma-1}\right]} = D_t e^{\left[\mu-\sigma^2\right](T-t)+\sigma^2\gamma(T-t)} \\ &= e^{\left[\mu-(1-\gamma)\sigma^2\right]T+\left[(1-\gamma)-\frac{1}{2}\right]\sigma^2 t+\sigma(z_t-z_0)} \end{aligned} \tag{15.45}$$

A simple application of Itô's lemma shows that equation (15.45) implies that the risky asset's price follows geometric Brownian motion:

$$dS_{r,t}/S_{r,t} = (1-\gamma)\,\sigma^2 dt + \sigma dz \tag{15.46}$$

Similarly, when all individuals are irrational, the deflated stock price, $S_{n,t}$, is

$$S_{n,t} = e^{\left[\mu-(1-\gamma-\eta)\sigma^2\right]T+\left[(1-\gamma-\eta)-\frac{1}{2}\right]\sigma^2 t+\sigma(z_t-z_0)} = S_{r,t}e^{\eta\sigma^2(T-t)} \tag{15.47}$$

and its rate of return follows the process

$$dS_{n,t}/S_{n,t} = (1-\gamma-\eta)\,\sigma^2 dt + \sigma dz \tag{15.48}$$

Note that in (15.47) and (15.48) the effect of η is similar to γ. When all individuals are irrational, if η is positive, the higher expected dividend growth acts like lower risk aversion in that individuals find the risky asset, relative to the zero-coupon bond, more attractive. Equation (15.47) shows that this greater demand raises the deflated stock price relative to that in an economy with all rational individuals, while equation (15.48) indicates that it also lowers the stock's equilibrium expected rate of return.

It is also interesting to note that (15.46) and (15.48) indicate that when the economy is populated by only one type of individual, the volatility of the risky asset's deflated return equals σ. In contrast, when both types of individuals populate the economy, the risky asset's volatility, $\sigma_{S,t}$, always exceeds σ. Applying Itô's lemma to (15.44), Kogan, Ross, Wang, and Westerfield prove that the risky asset's volatility satisfies the following bounds:[15]

$$\sigma \leq \sigma_{S,t} \leq \sigma\left(1+|\eta|\right) \tag{15.49}$$

The conclusion is that a diversity of beliefs has the effect of raising the equilibrium volatility of the risky asset.

For the special case in which rational and irrational individuals have logarithmic utility, that is, $\gamma = 0$, then (15.44) simplifies to

$$\begin{aligned} S_t &= \frac{1+E_t\left[\xi_T\right]}{E_t\left[\left(1+\xi_T\right)D_T^{-1}\right]} \\ &= D_t e^{\left[\mu-\sigma^2\right](T-t)} \frac{1+\xi_t}{1+\xi_t e^{-\eta\sigma^2(T-t)}} \\ &= e^{\left[\mu-\frac{1}{2}\sigma^2\right]T-\frac{1}{2}\sigma^2(T-t)+\sigma(z_t-z_0)} \frac{1+\xi_t}{1+\xi_t e^{-\eta\sigma^2(T-t)}} \end{aligned} \tag{15.50}$$

For this particular case, the risky asset's expected rate of return and variance, as a function of the distribution of wealth between the rational and irrational individuals, can be derived explicitly. Define

$$\alpha_t \equiv \frac{W_{r,t}}{W_{r,t}+W_{n,t}} = \frac{W_{r,t}}{S_t} \tag{15.51}$$

as the proportion of total wealth owned by the rational individuals. Using (15.40) and (15.44), we see that when $\gamma = 0$ this ratio equals

[15] The proof is given in Appendix B of (Kogan, Ross, Wang, and Westerfield 2006). Below, we show that this bound is satisfied for the case of individuals with logarithmic utility.

$$\alpha_t = \frac{E_t\left[\left(1+\varepsilon_{T,t}^{\frac{1}{1-\gamma}}\right)^{-\gamma} D_T^{\gamma}\right]}{E_t\left[\left(1+\varepsilon_{T,t}^{\frac{1}{1-\gamma}}\right)^{1-\gamma} D_T^{\gamma}\right]} = \frac{1}{1+E_t\left[\xi_T\right]} = \frac{1}{1+\xi_t} \tag{15.52}$$

Viewing S_t as a function of D_t and ξ_t as in the second line of (15.50), Itô's lemma can be applied to derive the mean and standard deviation of the risky asset's rate of return. The algebra is lengthy but results in the values

$$\sigma_{S,t} = \sigma + \eta\sigma\left[\frac{1}{1+e^{-\eta\sigma^2(T-t)}\left(\alpha_t^{-1}-1\right)} - \alpha_t\right] \tag{15.53}$$

and

$$\mu_{S,t} = \sigma_{S,t}^2 - \eta\sigma\left(1-\alpha_t\right)\sigma_{S,t} \tag{15.54}$$

where we have used $\alpha_t = 1/\left(1+\xi_t\right)$ to substitute out for ξ_t. Note that when $\alpha_t = 1$ or 0, equations (15.53) and (15.54) are consistent with (15.46) and (15.48) for the case of $\gamma = 0$.

Kogan, Ross, Wang, and Westerfield use their model to study how terminal wealth (consumption) is distributed between the rational and irrational individuals as the investment horizon, T, becomes large. The motivation for this comparative static exercise is the well-known conjecture made by Milton Friedman (Friedman 1953) that irrational traders cannot survive in a competitive market. The intuition is that when individuals trade based on the wrong beliefs, they will lose money to the rational traders, so that in the long run these irrational traders will deplete their wealth. Hence, in the long run, rational traders should control most of the economy's wealth and asset prices should reflect these rational individual's (correct) beliefs. The implication is that even when some individuals are irrational, markets should evolve toward long-run efficiency because irrational individuals will be driven to "extinction."

Kogan, Ross, Wang, and Westerfield introduce a definition of what would constitute the long-run dominance of rational individuals and, therefore, the relative extinction of irrational individuals. The *relative extinction* of an irrational individual would occur if

$$\lim_{T\to\infty} \frac{C_{n,T}}{C_{r,T}} = 0 \quad \text{a.s.} \tag{15.55}$$

which means that for arbitrarily small δ the probability of $\left| \lim_{T\to\infty} \frac{C_{n,T}}{C_{r,T}} \right| > \delta$ equals zero.[16] The relative extinction of a rational individual is defined symmetrically, and an individual is said to *survive relatively* in the long run if relative extinction does not occur.[17]

For the case of individuals having logarithmic utility, irrational individuals always suffer relative extinction. The proof of this is as follows. Rearranging (15.35), we have

$$\frac{C_{n,T}}{C_{r,T}} = (\lambda \xi_T)^{\frac{1}{1-\gamma}} \tag{15.56}$$

and for the case of $\gamma = 0$, (15.43) implies that $\lambda = 1$. Hence,

$$\begin{aligned} \frac{C_{n,T}}{C_{r,T}} &= \xi_T \\ &= e^{-\frac{1}{2}\sigma^2\eta^2 T + \sigma\eta(z_T - z_0)} \end{aligned} \tag{15.57}$$

Based on the strong law of large numbers for Brownian motions, it can be shown

[16] In general, a sequence of random variables, say, X_t, is said to converge to X almost surely (a.s.) if for arbitrary δ, the probability $P\left(\left|\lim_{t\longrightarrow\infty} X_t - X\right| > \delta\right) = 0$.

[17] One could also define the *absolute extinction* of the irrational individual. This would occur if $\lim_{T\to\infty} C_{n,T} = 0$ almost surely, and an individual is said to *survive absolutely* in the long run if absolute extinction does not occur. Relative survival is sufficient for absolute survival, but the converse is not true. Similarly, absolute extinction implies relative extinction, but the converse is not true.

that for any value of b

$$\lim_{T\to\infty} e^{aT+b(z_T-z_0)} = \begin{cases} 0 & a<0 \\ \infty & a>0 \end{cases} \tag{15.58}$$

where convergence occurs almost surely.[18] Since $-\frac{1}{2}\sigma^2\eta^2 < 0$ in (15.57), we see that equation (15.55) is proved.

The intuition for why irrational individuals become relatively extinct is due, in part, to the special properties of logarithmic utility. Note that the portfolio policy of the logarithmic rational individual is to maximize at each date t the utility

$$E_t\left[\ln C_{r,T}\right] = E_t\left[\ln W_{r,T}\right] \tag{15.59}$$

This is equivalent to maximizing the expected continuously compounded return per unit time:

$$E_t\left[\frac{1}{T-t}\ln\left(W_{r,T}/W_{r,t}\right)\right] = \frac{1}{T-t}\left[E_t\left[\ln\left(W_{r,T}\right)\right] - \ln\left(W_{r,t}\right)\right] \tag{15.60}$$

since $W_{r,t}$ is known at date t and $T-t>0$. Thus, from (15.60) the rational log utility individual follows a portfolio policy that maximizes $E_t\left[d\ln W_{r,t}\right]$ at each point in time. This portfolio policy is referred to as the "growth-optimum portfolio," because it maximizes the expected (continuously compounded) return on wealth.[19] Now given that in the model economy there is a single source of uncertainty affecting portfolio returns, dz, the processes for the rational and

[18] See section 2.9.A of (Karatzas and Shreve 1991).

[19] For the standard portfolio choice problem of selecting a portfolio from n risky assets and an instantaneously risk-free asset, we showed in equation (12.44) of Chapter 12 that the growth-optimum portfolio has the risky-asset portfolio weights $\omega_i^* = \sum_{j=1}^n \upsilon_{ij}\left(\mu_j - r\right)$. Note that this log utility investor's portfolio depends only on the current values of the investment opportunity set, and portfolio demands do not reflect a desire to hedge against changes in investment opportunities.

irrational individuals' wealths can be written as

$$dW_{r,t}/W_{r,t} = \mu_{r,t}dt + \sigma_{r,t}dz \tag{15.61}$$

$$dW_{n,t}/W_{n,t} = \mu_{n,t}dt + \sigma_{n,t}dz \tag{15.62}$$

where, in general, the expected rates of returns and volatilities, $\mu_{r,t}$, $\mu_{n,t}$, $\sigma_{r,t}$, and $\sigma_{n,t}$, are time varying. Applying Itô's lemma, it is straightforward to show that the process followed by the log of the ratio of the individuals' wealth is

$$\begin{aligned} d\ln\left(\frac{W_{n,t}}{W_{r,t}}\right) &= \left[\left(\mu_{n,t} - \frac{1}{2}\sigma_{n,t}^2\right) - \left(\mu_{r,t} - \frac{1}{2}\sigma_{r,t}^2\right)\right] dt + (\sigma_{n,t} - \sigma_{r,t})\, dz \\ &= E_t\left[d\ln W_{n,t}\right] - E_t\left[d\ln W_{r,t}\right] + (\sigma_{n,t} - \sigma_{r,t})\, dz \end{aligned} \tag{15.63}$$

Since the irrational individual chooses a portfolio policy that deviates from the growth-optimum portfolio, we know that $E_t\left[d\ln W_{n,t}\right] - E_t\left[d\ln W_{r,t}\right] < 0$, and thus $E_t\left[d\ln\left(W_{n,t}/W_{r,t}\right)\right] < 0$, making $d\ln\left(W_{n,t}/W_{r,t}\right)$ a process that is expected to steadily decline as $t \longrightarrow \infty$, which verifies Friedman's conjecture that irrational individuals lose wealth to rational ones in the long run.

While irrational individuals lose influence in the long run, as indicated by equations (15.50), (15.53), and (15.54), their presence may impact the level and dynamics of asset prices for substantial periods of time prior to becoming "extinct." Moreover, if as empirical evidence suggests, individuals have constant relative-risk-aversion utility with $\gamma < 0$ so that they are more risk averse than logarithmic utility, it turns out that Friedman's conjecture may not always hold. To see this, let us compute (15.56) for the general case of $\lambda = e^{-\gamma\eta\sigma^2 T}$:

$$\begin{aligned} \frac{C_{n,T}}{C_{r,T}} &= \left(\lambda\xi_T\right)^{\frac{1}{1-\gamma}} \\ &= e^{-\left[\gamma\eta + \frac{1}{2}\eta^2\right]\frac{\sigma^2}{1-\gamma}T + \frac{\sigma\eta}{1-\gamma}\left(z_T - z_0\right)} \end{aligned} \tag{15.64}$$

Thus, we see that the limiting behavior of $C_{n,T}/C_{r,T}$ is determined by the sign of the expression $\left[\gamma\eta + \frac{1}{2}\eta^2\right]$ or $\eta\left(\gamma + \frac{1}{2}\eta\right)$. Given that $\gamma < 0$, the strong law of large numbers allows us to conclude

$$\lim_{T\longrightarrow\infty}\frac{C_{n,T}}{C_{r,T}} = \begin{cases} 0 & \eta < 0 & \text{rational trader survives} \\ \infty & 0 < \eta < -2\gamma & \text{irrational trader survives} \\ 0 & -2\gamma < \eta & \text{rational trader survives} \end{cases} \tag{15.65}$$

When the irrational individual is pessimistic ($\eta < 0$) or strongly optimistic ($\eta > -2\gamma$), he becomes relatively extinct in the long run. However, when the irrational individual is moderately optimistic ($0 < \eta < -2\gamma$), the model has the opposite implication in that it is the rational individual who becomes relatively extinct in the long run. This parametric case is the reverse of Friedman's conjecture.

The intuition for these results comes from our previous discussion of a log utility investor's choice of the growth-optimal portfolio. When rational individuals are more risk averse than log utility ($\gamma < 0$), their demand for the risky asset is less than would be chosen by a log utility investor.[20] Ceteris paribus, the wealth of these $\gamma < 0$ investors would tend to grow more slowly than that of someone with log utility. When $\eta < 0$, irrationally pessimistic investors would demand even less of the risky asset than their rational counterparts, which would move them even farther away from the growth-optimal portfolio. Hence, in this case, a rational individual's wealth would tend to grow faster than the wealth of the irrational individual, so that the irrational individual would not survive in the long run.

When the irrational individual is optimistic ($\eta > 0$), her demand for the risky

[20] For example, recall from Chapter 12's analysis of the standard consumption-portfolio choice problem when investment opportunities are constant that equation (12.35), $\omega^* = \frac{\mu - r}{(1-\gamma)\sigma^2}$, implies that the demand for the risky asset decreases as risk aversion increases.

asset will exceed that of a rational investor. When her optimism is moderate, $(0 < \eta < -2\gamma)$, her portfolio demand is closer to the growth-optimal portfolio than is the portfolio demanded by the rational individual. Therefore, in this case, the moderately optimistic individual's wealth grows faster than that of the rational individual, so that the rational individual suffers relative extinction in the long run. In contrast, when the irrational individual is strongly optimistic $(\eta > -2\gamma)$, her demand for the risky asset is so great that her portfolio choice is farther from the growth-optimal portfolio than is the rational individual. For this case, the irrational individual's wealth tends to grow relatively slowly and, as in the pessimistic case, she does not survive in the long run.

The model outlined in this section is clearly a simplification of reality in that it assumes that individuals gain utility from only terminal, not interim, consumption. Interim consumption reduces the growth of wealth, and differences between rational and irrational individuals' consumption rates could affect their relative survivability. The model also assumes that rational and irrational individuals have the same preferences (levels of risk aversion). In general, an individual's portfolio choice, which affects his growth of wealth and survivability, is determined by risk aversion as well as beliefs. Hence, systematic differences between rational and irrational investors' risk aversions could influence the model's conclusions. In addition, one might expect that irrational individuals might learn over time of their mistakes since the historical distribution of the dividend process will tend to differ from their beliefs. The effect of such learning may be that irrationality could diminish with age.[21] Lastly, the model considers only one form of irrationality, namely, systematic optimism or pessimism. Other forms of irrationality have been identified that presumably

[21] However, there is empirical psychological evidence (Lord, Ross, and Lepper 1979) showing that individuals tend to persist too strongly in their initial beliefs after being exposed to contrary information.

would change the dynamics of wealth and of the equilibrium prices of risky assets.[22] Yet, the main conclusions of the model, that irrational investors may have a significant impact on asset prices and that they may not necessarily become extinct, are likely to remain robust.

15.3 Summary

There is a growing body of experimental and empirical research documenting that individuals do not always form beliefs rationally and do not always make decisions consistent with expected utility theory. Analyzing the asset pricing implications of such behavior is at an early stage. This chapter attempted to present two of the few general equilibrium models that incorporate psychological biases or irrationality. Interestingly, these models can be solved using techniques similar to those previously employed to derive models of rational, expected-utility-maximizing individuals. Both models in this chapter embed rationality as a special case, which makes it easy to see how their behavioral assumptions specifically affect the models' results.

Currently, there is no consensus among financial economists regarding the importance of incorporating aspects of behavioral finance into asset pricing theories. Some criticize behavioral finance theories as ad hoc explanations of anomalies that are not always mutually consistent. It is especially unclear whether a behavioral paradigm will be universally successful in supplanting asset pricing theories built on von Neumann-Morgenstern expected utility. However, it is likely that research exploring the asset pricing implications of behavioral biases will grow in coming years.

[22] Recent models incorporating various forms of irrationality (Barberis, Shleifer, and Vishny 1998); (Daniel, Hirshleifer, and Subrahmanyam 1998); and (Hong and Stein 1999) have been constructed to explain the empirical phenomena that stock returns display short-run positive serial correlation (momentum) and long-run negative serial correlation (reversals or mean reversion). See pages 1551-1556 of John Campbell's survey of asset pricing (Campbell 2000) for a summary of these and other behavioral finance models.

15.4 Exercises

1. In the Barberis, Huang, and Santos model, verify that the first-order conditions (15.16) and (15.17) lead to the envelope condition (15.18).

2. In the Barberis, Huang, and Santos model, solve for the price - dividend ratio, P_t/D_t, for Economy II when utility is standard constant relative risk aversion, that is,
$$E_0\left[\sum_{t=0}^{\infty}\delta^t\frac{C_t^\gamma}{\gamma}\right]$$

3. In the Kogan, Ross, Wang, and Westerfield model, verify that $\lambda = e^{-\gamma\eta\sigma^2 T}$ satisfies the equality $W_{r,0} = W_{n,0}$.

4. In the Kogan, Ross, Wang, and Westerfield model, suppose that both representative individuals are rational but have different levels of risk aversion. The first type of representative individual maximizes utility of the form
$$E_0\left[\frac{C_{r,T}^{\gamma_1}}{\gamma_1}\right]$$
and the second type of representative individual maximizes utility of the form
$$E_0\left[\frac{C_{n,T}^{\gamma_2}}{\gamma_2}\right]$$
where $1 > \gamma_1 > \gamma_2$. Assuming $W_{r,0} = W_{n,0}$, solve for the equilibrium price of the risky asset deflated by the discount bond maturing at date T.

Chapter 16

Asset Pricing with Differential Information

The asset pricing models in prior chapters assumed that individuals have common information. Now we will consider arguably more realistic situations where individuals can have different private information about an asset's future payoff or value. Because the literature on asset pricing in the presence of private information is vast, this chapter is meant to provide only a taste of this research area.[1] However, the two models that we present in this chapter, those of Sanford Grossman (Grossman 1976) and Albert "Pete" Kyle (Kyle 1985), are probably the two most common modeling frameworks in this field of research. Familiarity with these two models provides a segue to much additional theoretical research.

A topic of particular interest is the influence of private information on a risky asset's equilibrium price. We start by analyzing the Grossman model that shows

[1]More in-depth coverage of topics in this chapter includes books by Maureen O'Hara (O'Hara 1995) and Markus Brunnermeier (Brunnermeier 2001).

how individuals' information affects their demands for an asset and, via these demands, how private information is contained in the asset's equilibrium price. The model examines two equilibria: a "competitive," but not fully rational, equilibrium; and a fully-revealing rational expectations equilibrium.

Following this, we examine an extension of the Grossman model that includes an additional source of uncertainty, namely, shifts in the supply of the risky asset. A model of this type was developed in a number of studies, including (Grossman and Stiglitz 1980), (Hellwig 1980), (Diamond and Verrecchia 1981), and (Grundy and McNichols 1989). Importantly, in a rational expectations equilibrium this additional supply uncertainty makes the equilibrium asset price only partially reveal the private information of individuals.

We cover one additional model of a risky-asset market that also possesses an equilibrium where private information is partially revealed. It is Kyle's seminal market microstructure model. This model assumes a market for a particular security in which one agent, the so-called insider, has private information and trades with lesser-informed agents composed of a market maker and "noise" traders. The model solves for the strategic trading behavior of the insider and market maker and provides a theoretical framework for determining bid-ask spreads and the market impact of trades.

16.1 Equilibrium with Private Information

The model by Sanford Grossman (Grossman 1976) that we consider in this section examines how an investor's private information about a risky asset's future payoff affects her demand for that asset and, in turn, the asset's equilibrium price. In addition, it takes account of the idea that a rational individual can learn about others' private information from the risky asset's price, a concept known as "price discovery."

16.1.1 Grossman Model Assumptions

The Grossman model is based on the following assumptions.

Assets

This is a single-period portfolio choice problem. At the beginning of the period, traders can choose between a risk-free asset, which pays a known end-of-period return (1 plus the interest rate) of R_f, and a risky asset that has a beginning-of-period price of P_0 per share and an end-of-period random payoff (price) of $\widetilde{P}_1$ per share. The unconditional distribution of $\widetilde{P}_1$ is assumed to be normally distributed as $N(m, \sigma^2)$. The aggregate supply of shares of the risky asset is fixed at $\bar{X}$, but the risk-free asset is in perfectly elastic supply.

Trader Wealth and Preferences

There are n different traders. The i^{th} trader has beginning-of-period wealth W_{0i} and is assumed to maximize expected utility over end-of-period wealth, $\tilde{W}_{1i}$. Each trader is assumed to have constant absolute-risk-aversion (CARA) utility, but traders' levels of risk aversion are permitted to differ. Specifically, the form of the i^{th} trader's utility function is assumed to be

$$U_i(\tilde{W}_{1i}) \;=\; -e^{-a_i \tilde{W}_{1i}}, \;\; a_i > 0 \tag{16.1}$$

Trader Information

At the beginning of the period, the i^{th} trader observes y_i, which is a realized value from the noisy signal of the risky-asset end-of-period value

$$\tilde{y}_i \;=\; \tilde{P}_1 \;+\; \tilde{\epsilon}_i \tag{16.2}$$

where $\tilde{\epsilon}_i \sim N(0, \sigma_i^2)$ and is independent of $\tilde{P}_1$.

16.1.2 Individuals' Asset Demands

Let X_i be the number of shares of the risky asset chosen by the i^{th} trader at the beginning of the period. Thus, the i^{th} trader's wealth accumulation equation can be written as

$$\tilde{W}_{1i} \;=\; R_f W_{0i} \;+\; \left[\tilde{P}_1 - R_f P_0\right] X_i \tag{16.3}$$

Denote I_i as the information available to the i^{th} trader at the beginning of the period. The trader's maximization problem is then

$$\max_{X_i} E\left[U_i(\tilde{W}_{1i}) \mid I_i\right] \;=\; \max_{X_i} E\left[-e^{-a_i\left(R_f W_{0i} \;+\; \left[\tilde{P}_1 - R_f P_0\right] X_i\right)} \mid I_i\right] \tag{16.4}$$

Since $\tilde{W}_{1i}$ depends on $\tilde{P}_1$, it is normally distributed, and due to the exponential form of the utility function, (16.4) is the moment-generating function of a normal random variable. Therefore, as we have seen earlier in the context of mean-variance analysis, the maximization problem is equivalent to

$$\max_{X_i}\left\{E\left[\tilde{W}_{1i} \mid I_i\right] \;-\; \frac{1}{2}\, a_i \,\mathrm{Var}\left[\tilde{W}_{1i} \mid I_i\right]\right\} \tag{16.5}$$

or

$$\max_{X_i}\left\{X_i\left(E\left[\tilde{P}_1 \mid I_i\right] - R_f P_0\right) \;-\; \frac{1}{2}\, a_i\, X_i^2\, \mathrm{Var}\left[\tilde{P}_1 \mid I_i\right]\right\} \tag{16.6}$$

The first-order condition with respect to X_i then gives us the optimal number of shares held in the risky asset:

$$X_i \;=\; \frac{E\left[\tilde{P}_1 \mid I_i\right] - R_f\, P_0}{a_i\, \mathrm{Var}\left[\tilde{P}_1 \mid I_i\right]} \tag{16.7}$$

Equation (16.7) indicates that the demand for the risky asset is increasing in its expected excess return but declining in its price variance and the investor's risk aversion. Note that the CARA utility assumption results in the investor's demand for the risky asset being independent of wealth. This simplifies the derivation of the risky asset's equilibrium price.

16.1.3 A Competitive Equilibrium

Now consider an equilibrium in which each trader uses his knowledge of the unconditional distribution of $\tilde{P}_1$ along with the conditioning information from his private signal, y_i, so that $I_i = \{y_i\}$. Then using Bayes rule and the fact that $\tilde{P}_1$ and $\tilde{y}_i$ are jointly normally distributed with a squared correlation $\rho_i^2 \equiv \frac{\sigma^2}{\sigma^2 + \sigma_i^2}$, the i^{th} trader's conditional expected value and variance of $\tilde{P}_1$ are[2]

$$\begin{aligned} E\left[\tilde{P}_1 \mid I_i\right] &= m + \rho_i^2 \left(y_i - m\right) \\ \text{Var}\left[\tilde{P}_1 \mid I_i\right] &= \sigma^2 \left(1 - \rho_i^2\right) \end{aligned} \tag{16.8}$$

Substituting these into (16.7), we have

$$X_i = \frac{m + \rho_i^2 \left(y_i - m\right) - R_f \, P_0}{a_i \, \sigma^2 \left(1 - \rho_i^2\right)} \tag{16.9}$$

From the denominator of (16.9), one sees that the individual's demand for the risky asset is greater the lower his risk aversion, a_i, and the greater the precision of his signal (the closer is ρ_i to 1, that is, the lower is σ_i). Now by aggregating the individual traders' risky-asset demands for shares and setting the sum equal to the fixed supply of shares, we can solve for the equilibrium risky-asset price,

[2] A derivation of (16.8) is given as an end-of-chapter exercise. Note that ρ_i is the correlation coefficient since $\frac{\text{cov}(\tilde{P}_1, \tilde{y}_i)}{\sigma_{\tilde{P}_1} \sigma_{\tilde{y}_i}} = \frac{\sigma^2}{\sigma \sqrt{\sigma^2 + \sigma_i^2}} = \rho_i$.

P_0, that equates supply and demand:

$$\begin{aligned}\bar{X} &= \sum_{i=1}^{n}\left[\frac{m+\rho_i^2\,(y_i-m)-R_f\,P_0}{a_i\,\sigma^2\,(1-\rho_i^2)}\right] \\ &= \sum_{i=1}^{n}\left[\frac{m+\rho_i^2\,(y_i-m)}{a_i\,\sigma^2\,(1-\rho_i^2)}\right]-\sum_{i=1}^{n}\left[\frac{R_f\,P_0}{a_i\,\sigma^2\,(1-\rho_i^2)}\right]\end{aligned} \tag{16.10}$$

or

$$P_0 = \frac{1}{R_f}\left[\sum_{i=1}^{n}\frac{m+\rho_i^2\,(y_i-m)}{a_i\,\sigma^2\,(1-\rho_i^2)}-\bar{X}\right] \Bigg/ \left[\sum_{i=1}^{n}\frac{1}{a_i\,\sigma^2\,(1-\rho_i^2)}\right] \tag{16.11}$$

From (16.11) we see that the price reflects a weighted average of the traders' conditional expectations of the payoff of the risky asset. For example, the weight on the i^{th} trader's conditional expectation, $m+\rho_i^2(y_i-m)$, is

$$\frac{1}{a_i\,\sigma^2\,(1-\rho_i^2)} \Bigg/ \left[\sum_{i=1}^{n}\frac{1}{a_i\,\sigma^2\,(1-\rho_i^2)}\right] \tag{16.12}$$

The more precise (higher ρ_i) is trader i's signal or the lower is his risk aversion (lower a_i), the more aggressively he trades and, as a result, the more that the equilibrium price reflects his expectations.

16.1.4 A Rational Expectations Equilibrium

The solution for the price, P_0, in equation (16.11) can be interpreted as a competitive equilibrium: each trader uses information from his own signal and takes the price of the risky asset as given in formulating her demand for the risky asset. However, this equilibrium neglects the possibility that a trader might infer information about other traders' signals from the equilibrium price itself, what practitioners call "price discovery." In this sense, the previous equilibrium is not a rational expectations equilibrium. Why? Suppose traders initially formulate their demands according to equation (16.9), using only information about their own signals, and the price in (16.11) results. Then an individual trader could obtain information about the other traders' signals from the formula for P_0 in (16.11). Hence, this trader would have the incentive to change her demand from that initially formulated in (16.9). This implies that equation (16.11) would not be the rational expectations equilibrium price.

Therefore, to derive a fully rational expectations equilibrium, we need to allow traders' information sets to depend not only on their individual signals, but on the equilibrium price itself: $I_i = \{y_i, P_0^*(y)\}$ where $y \equiv (y_1 \;\; y_2 \;\; ... \;\; y_n)$ is a vector of the traders' individual signals and $P_0^*(y)$ is the rational expectations equilibrium price.[3]

In equilibrium, the aggregate demand for the shares of the risky asset must equal the aggregate supply, implying

$$\bar{X} \;=\; \sum_{i=1}^{n} \left[\frac{E\left[\tilde{P}_1 \mid y_i,\, P_0^*(y)\right] - R_f\, P_0^*(y)}{a_i \operatorname{Var}\left[\tilde{P}_1 \mid y_i,\, P_0^*(y)\right]} \right] \tag{16.13}$$

Now one can show that a rational expectations equilibrium exists when investors' signals have independent forecast errors and have equal accuracies. Specifically, it is assumed that in (16.2) the ϵ_i's are independent and have the same variance, $\sigma_i^2 = \sigma_\epsilon^2$, for $i = 1, ..., n$.

Theorem: There exists a rational expectations equilibrium with $P_0^*(y)$ given by

$$P_0^*(y) \;=\; \frac{1-\rho^2}{R_f} m + \frac{\rho^2}{R_f} \bar{y} - \frac{\sigma^2 \left(1-\rho^2\right)}{R_f \sum_{i=1}^{n} \frac{1}{a_i}} \bar{X} \tag{16.14}$$

where $\bar{y} \equiv \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} y_i$ and $\rho^2 \equiv \dfrac{\sigma^2}{\sigma^2 + \frac{\sigma_\epsilon^2}{n}}$.

[3] The theory of a rational expectations equilibrium was introduced by John F. Muth (Muth 1961). Robert E. Lucas won the 1995 Nobel prize in economics for developing and applying rational expectations theory in several papers, including (Lucas 1972), (Lucas 1976), and (Lucas 1987).

Proof: An intuitive outline of the proof is as follows.[4] Note that in (16.14), $P_0^*(y)$ is a linear function of $\bar{y}$ with a fixed coefficient of ρ^2/R_f. Therefore, if a trader observes $P_0^*(y)$ (and knows the structure of the model, that is, the other parameters), then he can invert this price formula to infer the value of $\bar{y}$. Now because all traders' signals were assumed to have equal precision (same σ_ϵ^2), the average signal, $\bar{y}$, is a sufficient statistic for the information contained in all of the other signals. Further, because of the assumed independence of the signals, the precision of this average of signals is proportional to the number of traders, n. Hence, the average signal would have the same precision as a single signal with variance $\frac{\sigma_\epsilon^2}{n}$.

Now if individual traders' demands are given by equation (16.9) but where y_i is replaced with $\bar{y}$ and ρ_i is replaced with ρ, then by aggregating these demands and setting them equal to $\bar{X}$ as in equation (16.10), we end up with the solution in equation (16.14), which is consistent with our initial assumption that traders can invert $P_0^*(y)$ to find $\bar{y}$. Hence, $P_0^*(y)$ in equation (16.14) is the rational expectations equilibrium price of the risky asset.

Note that the information, $\bar{y}$, reflected in the equilibrium price is superior to any single trader's private signal, y_i. In fact, since $\bar{y}$ is a sufficient statistic for all traders' information, it makes knowledge of any single signal, y_i, redundant. The equilibrium would be the same if all traders received the same signal, $\bar{y} \sim N(m, \sigma^2 + \frac{\sigma_\epsilon^2}{n})$ or if they all decided to share information on their private signals among each other before trading commenced.

Therefore, the above equilibrium is a *fully revealing* rational expectations equilibrium. The equilibrium price fully reveals all private information, a condition defined as strong-form market efficiency.[5] This result has some interesting features in that it shows that prices can aggregate relevant information

[4] See the original Grossman article (Grossman 1976) for details.

[5] This can be compared to semistrong form market efficiency where asset prices need only reflect all public information.

to help agents make more efficient investment decisions than would be the case if they relied solely on their private information and did not attempt to obtain information from the equilibrium price itself.

However, as shown by Sanford Grossman and Joseph Stiglitz (Grossman and Stiglitz 1980), this fully revealing equilibrium is not robust to some small changes in assumptions. Real-world markets are unlikely to be perfectly efficient. For example, suppose each trader needed to pay a tiny cost, c, to obtain his private signal, y_i. With any finite cost of obtaining information, the equilibrium would not exist, because each individual receives no additional benefit from knowing y_i given that they can observe $\bar{y}$ from the price. In other words, a given individual does not personally benefit from having private (inside) information in a fully revealing equilibrium. In order for individuals to benefit from obtaining (costly) information, we need an equilibrium where the price is only partially revealing. For this to happen, there needs to be one or more additional sources of uncertainty that add "noise" to individuals' signals, so that other agents cannot infer them perfectly. We now turn to an example of a noisy rational expectations equilibrium.

16.1.5 A Noisy Rational Expectations Equilibrium

Let us make the following changes to the Grossman model's assumptions along the lines of a model proposed by Bruce Grundy and Maureen McNichols (Grundy and McNichols 1989). Suppose that each trader begins the period with a random endowment of the risky asset. Specifically, trader i possesses ε_i shares of the risky asset so that her initial wealth is $W_{0i} = \varepsilon_i P_0$. The realization of ε_i is known only to trader i. Across all traders, the endowments, $\widetilde{\varepsilon}_i$, are independently and identically distributed with mean μ_X and variance $\sigma_X^2 n$. To simplify the problem, we assume that the number of traders is very large. If we define

$\widetilde{X}$ as the *per capita* supply of the risky asset and let n go to infinity, then by the Central Limit Theorem, $\widetilde{X}$ is a random variable distributed $N\left(\mu_X, \sigma_X^2\right)$. Note that in the limit as $n \to \infty$, the correlation between $\widetilde{\varepsilon}_i$ and $\widetilde{X}$ becomes zero, so that trader i's observation of her own endowment, $\widetilde{\varepsilon}_i$, provides no information about the per capita supply, $\widetilde{X}$.

Next, let us modify the type of signal received by each trader to allow for a common error as well as a trader-specific error. Trader i is assumed to receive the signal

$$\tilde{y}_i \;=\; \tilde{P}_1 \;+\; \widetilde{\omega} + \tilde{\epsilon}_i \tag{16.15}$$

where $\widetilde{\omega} \sim N(0,\, \sigma_\omega^2)$ is the common error independent of $\tilde{P}_1$ and, as before, the idiosyncratic error $\tilde{\epsilon}_i \sim N(0, \sigma_\epsilon^2)$ and is independent of $\tilde{P}_1$ and $\widetilde{\omega}$. Because of the infinite number of traders, it is realistic to allow for a common error so that traders, collectively, would not know the true payoff of the risky asset.

Recall from the Grossman model that the rational expectations equilibrium price in (16.14) was a linear function of $\overline{y}$ and $\overline{X}$. In the current model, the aggregate supply of the risky asset is not fixed, but random. However, this suggests that the equilibrium price will be of the form

$$P_0 = \alpha_0 + \alpha_1 \overline{y} + \alpha_2 \widetilde{X} \tag{16.16}$$

where now $\overline{y} \equiv \lim_{n\to\infty} \sum_i^n y_i/n = \tilde{P}_1 \;+\; \widetilde{\omega}$.

Although some assumptions differ, trader i's demand for the risky asset continues to be of the form in (16.7). Now recall that in a rational expectations equilibrium, investor i's information set includes not only her private information but also the equilibrium price: $I_i = \{y_i, P_0\}$. Given the assumed structure in (16.16) and the assumed normal distribution for $\widetilde{P}_1$, $\widetilde{X}$, and y_i, then investor

i optimally forecasts the end-of-period price as the projection

$$E\left[\widetilde{P}_1|I_i\right] = \beta_0 + \beta_1 P_0 + \beta_2 y_i \tag{16.17}$$

where

$$\begin{aligned}\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix} &= \begin{pmatrix}\alpha_1^2\left(\sigma^2+\sigma_\omega^2\right)+\alpha_2^2\sigma_X^2 & \alpha_1\left(\sigma^2+\sigma_\omega^2\right)\\ \alpha_1\left(\sigma^2+\sigma_\omega^2\right) & \sigma^2+\sigma_\omega^2+\sigma_\epsilon^2\end{pmatrix}^{-1}\begin{pmatrix}\alpha_1^2\sigma^2\\ \sigma^2\end{pmatrix}\\ \beta_0 &= m-\beta_1\left(\alpha_0-\alpha_1 m-\alpha_2\mu_X\right)-\beta_2 m\end{aligned} \tag{16.18}$$

If we then average the X_i in (16.7) over all investors, one obtains

$$\begin{aligned}X &= \frac{\beta_0+\left(\beta_1-R_f\right)P_0+\beta_2\overline{y}}{\overline{a}\text{Var}\left[\tilde{P}_1 \mid I_i\right]} \\ &= \frac{\beta_0}{\overline{a}\text{Var}\left[\tilde{P}_1 \mid I_i\right]}+\frac{\beta_1-R_f}{\overline{a}\text{Var}\left[\tilde{P}_1 \mid I_i\right]}P_0+\frac{\beta_2}{\overline{a}\text{Var}\left[\tilde{P}_1 \mid I_i\right]}\overline{y}\end{aligned} \tag{16.19}$$

where $\overline{a} \equiv 1/\left(\lim_{n\to\infty}\frac{1}{n}\sum_i^n\frac{1}{a_i}\right)$ is the harmonic mean of the investors' risk aversions. Now note that we can rewrite equation (16.16) as

$$X = -\frac{\alpha_0}{\alpha_2}+\frac{1}{\alpha_2}P_0-\frac{\alpha_1}{\alpha_2}\overline{y} \tag{16.20}$$

In a rational expectations equilibrium, the relationships between the variables X, P_0, and $\overline{y}$ must be consistent with the individual investors' expectations. This implies that the intercepts, and the coefficients on P_0 and on $\overline{y}$, must be identical in equations (16.19) and (16.20). By matching the intercepts and coefficients, we obtain three nonlinear equations in the three unknowns α_0, α_1, and α_2. Although explicit solutions for α_0, α_1, and α_2 cannot be obtained, we can still interpret some of the characteristics of the equilibrium. To see this, note that if the coefficients on $\overline{y}$ are equated, one obtains

$$-\frac{\alpha_1}{\alpha_2} = \frac{\beta_2}{\overline{a}\text{Var}\left[\tilde{P}_1 \mid I_i\right]} \tag{16.21}$$

Using (16.18) to substitute for β_2 and the variance of the projection of $\tilde{P}_1$ on I_i to substitute for $\text{Var}\left[\tilde{P}_1 \mid I_i\right]$, (16.21) can be rewritten as

$$-\frac{\alpha_1}{\alpha_2} = \frac{\sigma_X^2}{\overline{a}\left[\sigma_X^2\left(\sigma_\omega^2 + \sigma_\epsilon^2\right) + \left(\alpha_1/\alpha_2\right)^2\sigma_\omega^2\sigma_\epsilon^2\right]} \tag{16.22}$$

This is a cubic equation in α_1/α_2. The ratio α_1/α_2 is a measure of how aggressively an individual investor responds to his individual private signal, relative to the average signal, $\overline{y}$, reflected in P_0. To see this, note that if one uses (16.7), (16.19), (16.20), and $y_i - \overline{y} = \tilde{\epsilon}_i$, the individual's demand for the risky asset can be written as

$$X_i = \frac{\overline{a}}{a_i}\left(X - \frac{\alpha_1}{\alpha_2}\tilde{\epsilon}_i\right) \tag{16.23}$$

From (16.23) one sees that if there were no information differences, each investor would demand a share of the average supply of the risky asset, X, in proportion to the ratio of the harmonic average of risk aversions to his own risk aversion. However, unlike the fully revealing equilibrium of the previous section, the individual investor cannot perfectly invert the equilibrium price to find the average signal in (16.16) due to the uncertain aggregate supply shift, X. Hence, individual demands do respond to private information as reflected by $\tilde{\epsilon}_i$. The ratio α_1/α_2 reflects the simultaneous equation problem faced by the investor in trying to sort out a shift in supply, X, from a shift in aggregate demand generated by $\overline{y}$. From (16.22) note that if $\sigma_\omega^2 \to \infty$ or $\sigma_\epsilon^2 \to \infty$, so that investors' private signals become uninformative, then $\alpha_1/\alpha_2 \to 0$ and private information has no impact on demands or the equilibrium price. If, instead, $\sigma_\omega^2 = 0$, so that there is no common error, then (16.22) simplifies to

$$-\frac{\alpha_1}{\alpha_2} = \frac{1}{\overline{a}\sigma_\epsilon^2} \tag{16.24}$$

and (16.23) becomes

$$X_i = \frac{\overline{a}}{a_i}X - \frac{1}{a_i\sigma_\epsilon^2}\tilde{\epsilon}_i \tag{16.25}$$

so that an individual's demand responds to her private signal in direct proportion to the signal's precision and indirect proportion to her risk aversion.

A general insight of this noisy rational expectations model is that an investor forms her asset demand based on her private signal but also attempts to extract the private signals of other investors from the asset's equilibrium price. We now study another signal extraction problem but where the signal is reflected in the quantity of an asset being traded. The problem is one of a market maker who is charged with setting a competitive market price of an asset when some trades reflect private information.

16.2 Asymmetric Information, Trading, and Markets

Let us now consider another model with private information that is pertinent to a security market organized by a market maker. This market maker, who might be thought of as a specialist on a stock exchange or a security dealer in an over-the-counter market, sets a risky asset's price with the recognition that he may be trading at that price with a possibly better-informed individual. Albert "Pete" Kyle (Kyle 1985) developed this model, and it has been widely applied to study market microstructure issues. The model is similar to the previous one in that the equilibrium security price partially reveals the better-informed individual's private information. Also like the previous model, there is an additional source of uncertainty that prevents a fully revealing equilibrium, namely, orders from uninformed "noise," or "liquidity" traders who provide camouflage for the better-informed individual's insider trades. The model's results provide insights regarding the factors affecting bid-ask spreads and the market impact of trades.

16.2.1 Kyle Model Assumptions

The Kyle model is based on the following assumptions.

Asset Return Distribution

The model is a single-period model.[6] At the beginning of the period, agents trade in an asset that has a random end-of-period liquidation value of $\tilde{\nu} \sim N\left(p_0, \sigma_v^2\right)$.

Liquidity Traders

Noise traders have needs to trade that are exogenous to the model. It is assumed that they, as a group, submit a "market" order to buy $\tilde{u}$ shares of the asset, where $\tilde{u} \sim N\left(0, \sigma_u^2\right)$. $\tilde{u}$ and $\tilde{\nu}$ are assumed to be independently distributed.[7]

Better-Informed Traders

The single risk-neutral *insider* is assumed to have better information than the other agents. He knows with perfect certainty the realized end-of-period

---

[6] Kyle's paper (Kyle 1985) also contains a multiperiod continuous-time version of his single-period model. Jiang Wang (Wang 1993) has also constructed a continuous-time asset pricing model with asymmetrically informed investors who have constant absolute-risk-aversion utility.

[7] Why rational noise traders submit these orders has been modeled by assuming they have exogenous shocks to their wealth and need to rebalance their portfolio (Spiegel and Subrahmanyam 1992) or by assuming that they have uncertainty regarding the timing of their consumption (Gorton and Pennacchi 1993).

value of the risky security $\tilde{\nu}$ (but not $\tilde{u}$) and chooses to submit a market order of size x that maximizes his expected end-of-period profits.[8]

Competitive Market Maker

The single risk-neutral *market maker* (for example, a New York Stock Exchange specialist) receives the market orders submitted by the noise traders and the insider, which in total equal $\widetilde{u}+\widetilde{x}$. Importantly, the market maker cannot distinguish what part of this total order consists of orders made by noise traders and what part consists of the order of the insider. (The traders are anonymous.) The market maker sets the market price, p, and then takes the position $-(\widetilde{u}+\widetilde{x})$ to clear the market. It is assumed that market making is a perfectly competitive profession, so that the market maker sets the price p such that, given the total order submitted, his profit at the end of the period is expected to be zero.

16.2.2 Trading and Pricing Strategies

Since the noise traders' order is exogenous, we need only consider the optimal actions of the market maker and the insider.

The market maker observes only the total order flow, $u+x$. Given this information, he must then set the equilibrium market price p that gives him zero expected profits. Since his end-of-period profits are $-(\tilde{\nu}-p)(u+x)$, this implies that the price set by the market maker satisfies

$$p = E\left[\tilde{\nu} \mid u+x\right] \tag{16.26}$$

The information on the total order size is important to the market maker. The more positive the total order size, the more likely it is that x is large due to

[8] This assumption can be weakened to the case of the insider having uncertainty over $\tilde{\nu}$ but having more information on $\tilde{\nu}$ than the other traders. One can also allow the insider to submit "limit" orders, that is, orders that are a function of the equilibrium market price (a demand schedule), as in another model by Kyle (Kyle 1989).

the insider knowing that ν is greater than p_0. Thus, the market maker would tend to set p higher than otherwise. Similarly, the more negative is $u + x$, the more likely it is that x is low because the insider knows ν is below p_0 and is submitting a sell order. In this case, the market maker would tend to set p lower than otherwise. Thus, the *pricing rule* of the market maker is a function of $x + u$, that is, $P(x + u)$.

Since the insider sets x, it is an endogenous variable that depends on $\tilde{\nu}$. The insider chooses x to maximize his expected end-of-period profits, $\tilde{\pi}$, given knowledge of ν and the way that the market maker behaves in setting the equilibrium price:

$$\max_x E\left[\tilde{\pi} \mid \nu\right] = \max_x E\left[\left(\nu - P\left(x + \tilde{u}\right)\right) x \mid \nu\right] \tag{16.27}$$

An equilibrium in this model is a pricing rule chosen by the market maker and a trading strategy chosen by the insider such that 1) the insider maximizes expected profits, given the market maker's pricing rule; 2) the market maker sets the price to earn zero expected profits, given the trading strategy of the insider; and 3) the insider and market maker have rational expectations. That is, the equilibrium is a fixed point where each agent's actual behavior (e.g., pricing rule or trading strategy) is that which is expected by the other.

Insider's Trading Strategy

Suppose the market maker chooses a market price that is a linear function of the total order flow, $P(x + u) = \mu + \lambda(x + u)$. We will later argue that a linear pricing rule is optimal. If this is so, what is the insider's choice of x? From (16.27) we have

$$\max_x E\left[(\nu - P(x+\tilde{u}))\, x \mid \nu\right] = \max_x E\left[(\nu - \mu - \lambda(x+\tilde{u}))\, x \mid \nu\right] \quad (16.28)$$
$$= \max_x (\nu - \mu - \lambda x)\, x, \text{ since } E[\tilde{u}] = 0$$

Thus, the solution to the insider's problem in (16.28) is

$$x = \alpha + \beta\nu \quad (16.29)$$

where $\alpha = -\frac{\mu}{2\lambda}$ and $\beta = \frac{1}{2\lambda}$. Therefore, if the market maker uses a linear pricing rule, the optimal trading strategy for the insider is a linear trading rule.

Market Maker's Pricing Strategy

Next, let us return to the market maker's problem of choosing the market price that, conditional on knowing the total order flow, results in a competitive (zero) expected profit. Given the assumption that market making is a perfectly competitive profession, a market maker needs to choose the "best" possible estimate of $E[\tilde{\nu} \mid u + x]$ in setting the price $p = E[\tilde{\nu} \mid u + x]$. The maximum likelihood estimate of $E[\tilde{\nu} \mid u + x]$ is best in the sense that it attains maximum efficiency and is also the minimum-variance unbiased estimate.

Note that if the insider follows the optimal trading strategy, which according to equation (16.29) is $x = \alpha + \beta\tilde{\nu}$, then from the point of view of the market maker, $\tilde{\nu}$ and $y \equiv \tilde{u} + x = \tilde{u} + \alpha + \beta\tilde{\nu}$ are jointly normally distributed. Because ν and y are jointly normal, the maximum likelihood estimate of the mean of ν conditional on y is linear in y, that is, $E[\tilde{\nu} \mid y]$ is linear in y.[9] Hence, the previously assumed linear pricing rule is, in fact, optimal in equilibrium. Therefore, the market maker should use the maximum likelihood estimator,

[9] Earlier in this chapter, we saw an example of this linear relationship in equation (16.8).

which in the case of ν and y being normally distributed is equivalent to the "least squares" estimator. This estimator minimizes

$$\begin{aligned} E\left[(\tilde{\nu} - P(y))^2\right] &= E\left[(\tilde{\nu} - \mu - \lambda y)^2\right] \qquad (16.30) \\ &= E\left[(\tilde{\nu} - \mu - \lambda(\tilde{u} + \alpha + \beta\tilde{\nu}))^2\right] \end{aligned}$$

Thus, the optimal pricing rule equals $\mu + \lambda y$, where μ and λ minimize

$$\min_{\mu,\lambda} E\left[(\tilde{\nu}(1-\lambda\beta) - \lambda\tilde{u} - \mu - \lambda\alpha)^2\right] \qquad (16.31)$$

Recalling the assumptions $E[\nu] = p_0$, $E\left[(\nu - p_0)^2\right] = \sigma_v^2$, $E[u] = 0$, $E\left[u^2\right] = \sigma_u^2$, and $E[u\nu] = 0$, the objective function (16.31) can be written as

$$\min_{\mu,\lambda} (1-\lambda\beta)^2\left(\sigma_v^2 + p_0^2\right) + (\mu+\lambda\alpha)^2 + \lambda^2\sigma_u^2 - 2(\mu+\lambda\alpha)(1-\lambda\beta)p_0 \qquad (16.32)$$

The first-order conditions with respect to μ and λ are

$$\mu = -\lambda\alpha + p_0(1-\lambda\beta) \qquad (16.33)$$

$$\begin{aligned} 0 &= -2\beta(1-\lambda\beta)\left(\sigma_v^2 + p_0^2\right) + 2\alpha(\mu+\lambda\alpha) + 2\lambda\sigma_u^2 \\ &\quad -2p_0\left[-\beta(\mu+\lambda\alpha) + \alpha(1-\lambda\beta)\right] \qquad (16.34) \end{aligned}$$

Substituting $\mu + \lambda\alpha = p_0(1-\lambda\beta)$ from (16.33) into (16.34), we see that (16.34) simplifies to

$$\lambda = \frac{\beta\sigma_v^2}{\beta^2\sigma_v^2 + \sigma_u^2} \tag{16.35}$$

Substituting in for the definitions $\alpha = -\frac{\mu}{2\lambda}$ and $\beta = \frac{1}{2\lambda}$ in (16.33) and (16.35), we have

$$\mu = p_0 \tag{16.36}$$

$$\lambda = \tfrac{1}{2}\frac{\sigma_v}{\sigma_u} \tag{16.37}$$

In summary, the equilibrium price is

$$p = p_0 + \tfrac{1}{2}\frac{\sigma_v}{\sigma_u}\left(\tilde{u} + \tilde{x}\right) \tag{16.38}$$

where the equilibrium order submitted by the insider is

$$x = \frac{\sigma_u}{\sigma_v}\left(\tilde{\nu} - p_0\right) \tag{16.39}$$

16.2.3 Analysis of the Results

From (16.39), we see that the greater the volatility (amount) of noise trading, σ_u, the larger is the magnitude of the order submitted by the insider for a given deviation of ν from its unconditional mean. Hence, the insider trades more actively on his private information the greater the "camouflage" provided by noise trading. Greater noise trading makes it more difficult for the market maker to extract the "signal" of insider trading from the noise. Note that if equation (16.39) is substituted into (16.38), one obtains

$$\begin{aligned} p &= p_0 + \tfrac{1}{2}\frac{\sigma_v}{\sigma_u}\tilde{u} + \tfrac{1}{2}\left(\tilde{\nu} - p_0\right) \\ &= \tfrac{1}{2}\left(\frac{\sigma_v}{\sigma_u}\tilde{u} + p_0 + \tilde{\nu}\right) \end{aligned} \tag{16.40}$$

Thus, we see that only one-half of the insider's private information, $\frac{1}{2}\ \tilde{\nu}$, is reflected in the equilibrium price, so that the price is *not fully revealing.*[10] To obtain an equilibrium of incomplete revelation of private information, it is necessary to have a second source of uncertainty, namely, the amount of noise trading.

Using (16.39) and (16.40), we can calculate the insider's expected profits:

$$E\left[\tilde{\pi}\right] = E\left[x\left(\nu - p\right)\right] = E\left[\frac{\sigma_u}{\sigma_v}\left(\tilde{\nu} - p_0\right)\tfrac{1}{2}\left(\nu - p_0 - \frac{\sigma_v}{\sigma_u}\tilde{u}\right)\right] \tag{16.41}$$

Conditional on knowing ν, that is, after learning the realization of ν at the beginning of the period, the insider expects profits of

$$E\left[\tilde{\pi} \mid \nu\right] = \tfrac{1}{2}\frac{\sigma_u}{\sigma_v}\left(\nu - p_0\right)^2 \tag{16.42}$$

Hence, the larger ν's deviation from p_0, the larger the expected profit. Unconditional on knowing $\tilde{\nu}$, that is, before the start of the period, the insider expects a profit of

$$E\left[\tilde{\pi}\right] = \tfrac{1}{2}\frac{\sigma_u}{\sigma_v}E\left[\left(\tilde{\nu} - p_0\right)^2\right] = \tfrac{1}{2}\sigma_u\sigma_v \tag{16.43}$$

which is proportional to the standard deviations of noise traders' orders and the end-of-period value of ν.

[10] A fully revealing price would be $p = \tilde{\nu}$.

Since, by assumption, the market maker sets the security price in a way that gives him zero expected profits, the expected profits of the insider equals the expected losses of the noise traders. In other words, it is the noise traders that lose, on average, from the presence of the insider. Due to the market maker's inability to distinguish between informed (insider) and uninformed (noise trader) orders, they are treated the same under his pricing rule. Thus, on average, noise traders' buy (*sell*) orders are executed at a higher (*lower*) price than p_0.

From equation (16.38), we see that $\lambda = \frac{1}{2}\frac{\sigma_v}{\sigma_u}$ is the amount that the market maker raises the price when the total order flow, $(u + x)$, goes up by 1 unit.[11] This can be thought of as relating to the security's bid-ask spread, that is, the difference in the price for sell orders versus buy orders, although here sell and buy prices are not fixed but are a function of the order size since the pricing rule is linear. Moreover, since the amount of order flow necessary to raise the price by \$1 equals $1/\lambda = 2\frac{\sigma_u}{\sigma_v}$, the model provides a measure of the "depth" of the market, or market "liquidity." The higher is the proportion of noise trading to the value of insider information, $\frac{\sigma_u}{\sigma_v}$, the deeper, or more liquid, is the market.

Intuitively, the more noise traders relative to the value of insider information, the less the market maker needs to adjust the price in response to a given order, since the likelihood of the order being that of a noise trader, rather than an insider, is greater. The more noise traders there are (that is, the greater is σ_u), the greater is the expected profit of the insider (see equation (16.43)) and the greater is the *total* expected loss of the noise traders. However, the expected loss per *individual* noise trader falls with the greater level of noise trading.[12]

[11] It is now common in the market microstructure literature to refer to this measure of order flow and liquidity as "Kyle's lambda."

[12] Gary Gorton and George Pennacchi (Gorton and Pennacchi 1993) derive this result by modeling individual liquidity traders.

16.3 Summary

The models considered in this chapter analyze the degree to which private information about an asset's future payoff or value is reflected in the asset's current price. An investor's private information affects an asset's price by determining the investor's desired demand (long or short position) for the asset, though the investor's demand also is tempered by risk aversion. More subtly, we saw that a rational investor can also learn about the private information of other investors through the asset's price itself, and this price discovery affects the investors' equilibrium demands. Indeed, under some circumstances, the asset's price may fully reveal all relevant private information such that any individual's private information becomes redundant.

Perhaps more realistically, there are non-information-based factors that affect the net supply or demand for an asset. These "noise" factors prevent investors from perfectly inferring the private information signals of others, resulting in an asset price that is less than fully revealing. Noise provides camouflage for investors with private information, allowing these traders to profit from possessing such information. Their profits come at the expense of liquidity traders since the greater the likelihood of private information regarding a security, the larger will be the security's bid-ask spread. Hence, this theory predicts that a security's liquidity is determined by the degree of noise (non-information-based) trading relative to insider (private-information-based) trading.

16.4 Exercises

1. Show that the maximization problem in objective function (16.6) is equivalent to the maximization problem in (16.4).

2. Show that the results in (16.8) can be derived from Bayes rule and the assumption that $\tilde{P}_1$ and $\tilde{y}_i$ are normally distributed.

3. Consider a special case of the Grossman model. Traders can choose between holding a risk-free asset, which pays an end-of-period return of R_f, and a risky asset that has a beginning-of-period price of P_0 per share and an end-of-period payoff (price) of $\widetilde{P}_1$ per share. The unconditional distribution of $\widetilde{P}_1$ is assumed to be $N\left(m, \sigma^2\right)$. The risky asset is assumed to be a derivative security, such as a futures contract, so that its *net supply equals zero*.

   There are two different traders who maximize expected utility over end-of-period wealth, $\widetilde{W}_{1i}$, $i = 1, 2$. The form of the i^{th} trader's utility function is

$$U_i\left(\widetilde{W}_{1i}\right) = -e^{-a_i \widetilde{W}_{1i}},\ a_i > 0$$

   At the beginning of the period, the i^{th} trader observes y_i, which is a noisy signal of the end-of-period value of the risky asset

$$y_i = \widetilde{P}_1 + \widetilde{\epsilon}_i$$

   where $\epsilon_i \sim N\left(0, \sigma_\epsilon^2\right)$ and is independent of $\widetilde{P}_1$. Note that the variances of the traders' signals are the same. Also assume $E\left[\epsilon_1 \epsilon_2\right] = 0$.

a. Suppose each trader does not attempt to infer the other trader's information from the equilibrium price, P_0. Solve for each of the traders' demands for the risky asset and the equilibrium price, P_0.

b. Now suppose each trader does attempt to infer the other's signal from the equilibrium price, P_0. What will be the rational expectations equilibrium price in this situation? What will be each of the traders' equilibrium demands for the risky asset?

4. In the Kyle model (Kyle 1985), replace the original assumption **Better-Informed Traders** with the following new one:

   The single risk-neutral *insider* is assumed to have better information than the other agents. He observes a signal of the asset's end-of-period value equal to

$$s = \widetilde{v} + \widetilde{\varepsilon}$$

   where $\widetilde{\varepsilon} \sim N\left(0, \sigma_s^2\right)$ and $\widetilde{\varepsilon}$ is distributed independently of $\tilde{u}$ and $\tilde{\nu}$. The insider does not observe $\tilde{u}$ but chooses to submit a market order of size x that maximizes his expected end-of-period profits.

a. Suppose that the market maker's optimal price-setting rule is a linear function of the order flow

$$p = \mu + \lambda\left(u + x\right)$$

   Write down the expression for the insider's expected profits given this pricing rule.

b. Take the first-order condition with respect to x and solve for the insider's optimal trading strategy as a function of the signal and the parameters of the market maker's pricing rule.

c. Given the form of the insider's optimal trading strategy in the previous question, solve for the parameters μ and λ of the market maker's optimal price-setting rule $p = \mu + \lambda(u + x)$. How does the response of the price to a unit change in the order flow, λ, vary with the insider's signal error variance, σ_s^2?

5. Consider a variation of the Kyle model (Kyle 1985). Replace the orginal assumption **Liquidity Traders** with the following new one:

Noise traders have needs to trade that are exogenous to the model. It is assumed that they, as a group, submit a "market" order to buy $\tilde{u}$ shares of the asset, where $\tilde{u} \sim N(0, \sigma_u^2)$. $\tilde{u}$ and $\tilde{\nu}$ are assumed to be correlated with correlation coefficient ρ.

Note that the only change is that, instead of the original Kyle model's assumption that $\tilde{u}$ and $\tilde{\nu}$ are uncorrelated, they are now assumed to have nonzero correlation coefficient ρ.

a. Suppose that the market maker's optimal price-setting rule is a linear function of the order flow

$$p = \mu + \lambda(u + x)$$

Write down the expression for the insider's expected profits given this pricing rule. Hint: to find the conditional expectation of $\widetilde{u}$, it might be helpful to write it as a weighted average of $\widetilde{v}$ and another normal random variable uncorrelated with $\widetilde{v}$.

b. Take the first-order condition with respect to x and solve for the insider's optimal trading strategy as a function of v and the parameters of the market maker's pricing rule.

c. For a given pricing rule (given μ and λ) and a realization of $v > p_0$, does the insider trade more or less when $\rho > 0$ compared to the case of $\rho = 0$? What is the intuition for this result? How might a positive value for ρ be interpreted as some of the liquidity traders being better-informed traders? What insights might this result have for a market with multiple insiders (informed traders)?

Chapter 17

Models of the Term Structure of Interest Rates

This chapter provides an introduction to the main approaches for modeling the term structure of interest rates and for valuing fixed-income derivatives. It is not meant to be a comprehensive review of this subject. The literature on term structure models is voluminous, and many surveys on this topic, including (Dai and Singleton 2004), (Dai and Singleton 2003), (Maes 2003), (Piazzesi 2005a), (Rebonato 2004), and (Yan 2001), have appeared in recent years. The more modest objective of this chapter is to outline the major theories for valuing default-free bonds and bond derivatives, such as Treasury bills, notes, bonds, and their derivatives. The next chapter analyzes the valuation of default-risky bonds.

This chapter is comprised of two main sections. The first discusses models used to derive the equilibrium bond prices of different maturities in terms of particular state variables. One way to think about these models is that the state variables are the models' "input," while the values of different maturity bonds are the models' "output." The second section covers models that value fixed-income derivatives, such as interest rate caps and swaptions, in terms of a given maturity structure of bond prices. In contrast, these models take the term structure of observed bond prices as the input and have derivative values as the models' output.

17.1 Equilibrium Term Structure Models

Equilibrium term structure models describe the prices (or, equivalently, the yields) of different maturity bonds as functions of one or more state variables or "factors." The Vasicek model (Vasicek 1977), introduced in Chapter 9 (see equation 9.41), and the Cox, Ingersoll, and Ross model (Cox, Ingersoll, and Ross 1985b), presented in Chapter 13 (see equation 13.51), were examples of single-factor models. The single factor in the Vasicek model was the instantaneous-maturity interest rate, denoted $r(t)$, which was assumed to follow the Ornstein-Uhlenbeck process (9.30). In Cox, Ingersoll, and Ross's one-factor model, the factor was a variable that determined the expected returns of the economy's production processes. In equilibrium, the instantaneous-maturity interest rate was proportional to this factor and inherited its dynamics. This interest rate followed the square root process in equation (13.49).

Empirical evidence finds that term structure movements are driven by multiple factors.[1] In many multifactor models, the factors are latent (unobserved) variables that are identified by data on the yields of different maturity bonds. Recently, however, economists have renewed their interest in models that link term structure factors with observed macroeconomic variables.[2] A motivation

[1] For example, a principal components analysis by Robert Litterman and Jose Scheinkman (Litterman and Scheinkman 1988) finds that at least three factors are required to describe U.S. Treasury security movements. They relate these factors to the term structure's level, slope, and curvature.

[2] Francis Diebold, Monika Piazzesi, and Glenn Rudebusch (Diebold, Piazzesi, and

for these models is to better understand the relationship between the term structure of interest rates and the macroeconomy, with the potential of using term structure movements to forecast macroeconomic cycles.

Given the importance of multiple factors in term structure dynamics, let us generalize the pricing relationships for default-free, zero-coupon bonds that we developed in earlier chapters. We consider a situation where multiple factors determine bond prices and assume that there are n state variables, x_i, $i = 1, ..., n$, that follow the multivariate diffusion process

$$d\mathbf{x} = \mathbf{a}\left(t, \mathbf{x}\right) dt + \mathbf{b}\left(t, \mathbf{x}\right) \mathbf{dz} \tag{17.1}$$

where $\mathbf{x} = (x_1 ... x_n)'$; $\mathbf{a}(t, \mathbf{x})$ is an $n \times 1$ vector; $\mathbf{b}(t, \mathbf{x})$ is an $n \times n$ matrix; and $\mathbf{dz} = (dz_1 ... dz_n)'$ is an $n \times 1$ vector of independent Brownian motion processes so that $dz_i dz_j = 0$ for $i \neq j$.[3] This specification permits any general correlation structure for the state variables. Note that the instantaneous covariance matrix of the state variables is given by $\mathbf{b}(t, \mathbf{x}) \mathbf{b}(t, \mathbf{x})'$.

Define $P(t, T, \mathbf{x})$ as the date t price of a default-free, zero-coupon bond that pays 1 at date T. Itô's lemma gives the process followed by this bond's price:

$$dP\left(t, T, \mathbf{x}\right) / P\left(t, T, \mathbf{x}\right) = \mu_p\left(t, T, \mathbf{x}\right) dt + \boldsymbol{\sigma}_p\left(t, T, \mathbf{x}\right)' \mathbf{dz} \tag{17.2}$$

where the bond's expected rate of return equals

$$\mu_p\left(t, T, \mathbf{x}\right) = \left(\mathbf{a}\left(t, \mathbf{x}\right)' \mathbf{P}_x + P_t + \tfrac{1}{2}\text{Trace}\left[\mathbf{b}\left(t, \mathbf{x}\right) \mathbf{b}\left(t, \mathbf{x}\right)' \mathbf{P}_{xx}\right]\right) / P\left(t, T, \mathbf{x}\right) \tag{17.3}$$

Rudebusch 2005) discuss empirical estimation of term structure models using macroeconomic factors. An example of this approach is given by Andrew Ang and Monika Piazzesi (Ang and Piazzesi 2003).

[3] As discussed in Chapter 10, the independence assumption is not important. If there are correlated sources of risk (Brownian motions), they can be redefined by a linear transformation to be represented by n orthogonal risk sources.

and $\boldsymbol{\sigma}_p(t,T,\mathbf{x})$ is an $n \times 1$ vector of the bond's volatilities equal to

$$\boldsymbol{\sigma}_p(t,T,\mathbf{x}) = \mathbf{b}(t,\mathbf{x})' \mathbf{P}_x / P(t,T,\mathbf{x}) \tag{17.4}$$

and where $\mathbf{P}_x$ is an $n \times 1$ vector whose i^{th} element equals the partial derivative P_{x_i}; $\mathbf{P}_{xx}$ is an $n \times n$ matrix whose i,j^{th} element is the second-order partial derivative $P_{x_i x_j}$; and Trace$[A]$ is the sum of the diagonal elements of a square matrix A.

Similar to the Black-Scholes hedging argument discussed in Chapter 9 and applied to derive the Vasicek model, we can form a hedge portfolio of $n+1$ bonds having distinctly different maturities. By appropriately choosing the portfolio weights for these $n+1$ bonds, the n sources of risk can be hedged so that the portfolio generates a riskless return. In the absence of arbitrage, this portfolio's return must equal the riskless rate, $r(t,\mathbf{x})$. Making this no-arbitrage restriction produces the implication that each bond's expected rate of return must satisfy

$$\mu_p(t,T,\mathbf{x}) = r(t,\mathbf{x}) + \boldsymbol{\Theta}(t,\mathbf{x})' \boldsymbol{\sigma}_p(t,T,\mathbf{x}) \tag{17.5}$$

where $\boldsymbol{\Theta}(t,\mathbf{x}) = (\theta_1 ... \theta_n)'$ is the $n \times 1$ vector of market prices of risks associated with each of the Brownian motions in $\mathbf{dz} = (dz_1 ... dz_n)'$. By equating (17.5) to the process for $\mu_p(t,T,\mathbf{x})$ given by Itô's lemma in (17.3), we obtain the equilibrium partial differential equation (PDE)

$$\tfrac{1}{2}\text{Trace}\left[\mathbf{b}(t,\mathbf{x})\,\mathbf{b}(t,\mathbf{x})' \mathbf{P}_{xx}\right] + \left[\mathbf{a}(t,\mathbf{x}) - \mathbf{b}(t,\mathbf{x})\,\boldsymbol{\Theta}\right]' \mathbf{P}_x - rP + P_t = 0 \tag{17.6}$$

Given functional forms for $\mathbf{a}(t,\mathbf{x})$, $\mathbf{b}(t,\mathbf{x})$, $\boldsymbol{\Theta}(t,\mathbf{x})$, $r(t,\mathbf{x})$, this PDE can be solved subject to the boundary condition $P(T,T,\mathbf{x}) = 1$.

Note that equation (17.6) depends on the expected changes in the factors under the risk-neutral measure Q, $\mathbf{a}(t,\mathbf{x}) - \mathbf{b}(t,\mathbf{x})\,\boldsymbol{\Theta}$, rather than the factors' expected changes under the physical measure P, $\mathbf{a}(t, \mathbf{x})$. Hence, to price bonds, one could simply specify only the factors' risk-neutral processes.[4] This insight is not surprising, because we saw in Chapter 10 that the Feynman-Kac solution to this PDE is the risk-neutral pricing equation (10.61):

$$P(t, T, \mathbf{x}) = \widehat{E}_t \left[e^{-\int_t^T r(s,\mathbf{x})ds} \times 1 \right] \tag{17.7}$$

In addition to the pricing relations (17.6) and (17.7), we saw that a third pricing approach can be based on the pricing kernel that follows the process

$$dM/M = -r(t, \mathbf{x})\, dt - \mathbf{\Theta}(t, \mathbf{x})' \, \mathbf{dz} \tag{17.8}$$

In this case, pricing can be accomplished under the physical measure based on the formula

$$P(t, T, \mathbf{x}) = E_t \left[\frac{M(T)}{M(t)} \times 1 \right] \tag{17.9}$$

Thus far, we have placed few restrictions on the factors and their relationship to the short rate, $r(t, \mathbf{x})$, other than to assume that the factors follow the Markov diffusion processes (17.1). Let us next consider some popular parametric forms.

17.1.1 Affine Models

We start with models in which the yields of zero-coupon bonds are linear or "affine" functions of state variables. This class of models includes those of Oldrich Vasicek (Vasicek 1977) and John Cox, Jonathan Ingersoll, and Stephen Ross (Cox, Ingersoll, and Ross 1985b). Affine models are attractive because they lead to bond price formulas that are relatively easy to compute and because the parameters of the state variable processes can often be estimated using

[4] However, if the factors are observable variables for which data are available, it may be necessary to specify their physical processes if empirical implementations of the model require estimates for $\mathbf{a}(\mathbf{x}, t)$ and $\mathbf{b}(\mathbf{x}, t)$.

relatively straightforward econometric techniques.

Recall that a zero-coupon bond's continuously compounded yield, $Y(t,T,\mathbf{x})$, is defined from its price by the relation

$$P(t,T,\mathbf{x}) = e^{-Y(t,T,\mathbf{x})(T-t)} \tag{17.10}$$

One popular class of models assumes that zero-coupon bonds' continuously compounded yields are affine functions of the factors. Defining the time until maturity as $\tau \equiv T-t$, this assumption can be written as

$$Y(t,T,\mathbf{x})\,\tau = A(\tau) + \mathbf{B}(\tau)'\,\mathbf{x} \tag{17.11}$$

where $A(\tau)$ is a scalar function and $\mathbf{B}(\tau)$ is an $n \times 1$ vector of functions that do not depend on the factors, $\mathbf{x}$. Because at maturity $P(T,T,\mathbf{x}) = 1$, equation (17.11) implies that $A(0) = 0$ and $\mathbf{B}(0)$ is an $n \times 1$ vector of zeros. Another implication of (17.11) is that the short rate is also affine in the factors since

$$r(t,\mathbf{x}) = \lim_{T \to t} Y(t,T,\mathbf{x}) = \lim_{\tau \to 0} \frac{A(\tau) + \mathbf{B}(\tau)'\,\mathbf{x}}{\tau} \tag{17.12}$$

so that we can write $r(t,\mathbf{x}) = \alpha + \boldsymbol{\beta}'\mathbf{x}$, where $\alpha = \partial A(0)/\partial\tau$ is a scalar and $\boldsymbol{\beta} = \partial \mathbf{B}(0)/\partial\tau$ is an $n \times 1$ vector of constants.

Under what conditions regarding the factors' dynamics would the no-arbitrage, equilibrium bond yields be affine in the state variables? To answer this, let us substitute the affine yield assumption of (17.10) and (17.11) into the general no-arbitrage PDE of (17.6). Doing so, one obtains

$$\begin{aligned} &\tfrac{1}{2}\mathbf{B}(\tau)'\,\mathbf{b}(t,\mathbf{x})\,\mathbf{b}(t,\mathbf{x})'\,\mathbf{B}(\tau) - [\mathbf{a}(t,\mathbf{x}) - \mathbf{b}(t,\mathbf{x})\,\boldsymbol{\Theta}]'\,\mathbf{B}(\tau) \\ &+ \tfrac{\partial A(\tau)}{\partial\tau} + \tfrac{\partial \mathbf{B}(\tau)'}{\partial\tau}\mathbf{x} = \alpha + \boldsymbol{\beta}'\mathbf{x} \end{aligned} \tag{17.13}$$

Darrell Duffie and Rui Kan (Duffie and Kan 1996) characterize sufficient conditions for a solution to equation (17.13). Specifically, two of the conditions are that the factors' risk-neutral instantaneous expected changes and variances are affine in $\mathbf{x}$. In other words, if the state variables' risk-neutral drifts and variances are affine in the state variables, so are the equilibrium bond price yields. These conditions can be written as

$$\mathbf{a}(t,\mathbf{x}) - \mathbf{b}(t,\mathbf{x})\,\boldsymbol{\Theta} = \boldsymbol{\kappa}\,(\overline{\mathbf{x}} - \mathbf{x}) \tag{17.14}$$

$$\mathbf{b}(t,\mathbf{x}) = \boldsymbol{\Sigma}\sqrt{\mathbf{s}(\mathbf{x})} \tag{17.15}$$

where $\overline{\mathbf{x}}$ is an $n \times 1$ vector of constants, $\boldsymbol{\kappa}$ and $\boldsymbol{\Sigma}$ are $n \times n$ matrices of constants, and $\mathbf{s}(\mathbf{x})$ is an $n \times n$ diagonal matrix with the i^{th} diagonal term

$$\mathbf{s}_i(\mathbf{x}) = s_{oi} + \mathbf{s}_{1i}'\mathbf{x} \tag{17.16}$$

where s_{oi} is a scalar constant and $\mathbf{s}_{1i}$ is an $n \times 1$ vector of constants. Now, because the state variables' covariance matrix equals $\mathbf{b}(t,\mathbf{x})\,\mathbf{b}(t,\mathbf{x})' = \boldsymbol{\Sigma}\mathbf{s}(\mathbf{x})\,\boldsymbol{\Sigma}'$, additional conditions are needed to ensure that this covariance matrix remains positive definite for all possible realizations of the state variable, $\mathbf{x}$. Qiang Dai and Kenneth Singleton (Dai and Singleton 2000) and Darrell Duffie, Damir Filipovic, and Walter Schachermayer (Duffie, Filipovic, and Schachermayer 2002) derive these conditions.[5]

Given (17.14), (17.15), and (17.16), the partial differential equation in (17.13) can be rewritten as

[5] These conditions can have important consequences regarding the correlation between the state variables. For example, if the state variables follow a multivariate Ornstein-Uhlenbeck process, so that the model is a multifactor extension of the Vasicek model given in (9.41), (9.42), and (9.43), then any general correlation structure between the state variables is permitted. Terence Langetieg (Langeteig 1980) has analyzed this model. However, if the state variables follow a multivariate square root process, so that the model is a multifactor extension of the Cox, Ingersoll, and Ross model given in (13.51), (13.52), and (13.53), then the correlation between the state variables must be nonnegative.

$$\begin{aligned} & \tfrac{1}{2}\mathbf{B}(\tau)'\,\boldsymbol{\Sigma}\mathbf{s}(\mathbf{x})\,\boldsymbol{\Sigma}'\mathbf{B}(\tau) - [\boldsymbol{\kappa}(\overline{\mathbf{x}}-\mathbf{x})]'\,\mathbf{B}(\tau) + \frac{\partial A(\tau)}{\partial\tau} + \frac{\partial \mathbf{B}(\tau)'}{\partial\tau}\mathbf{x} \\ = \; & \alpha + \boldsymbol{\beta}'\mathbf{x} \end{aligned} \tag{17.17}$$

Note that this equation is linear in the state variables, $\mathbf{x}$. For the equation to hold for all values of $\mathbf{x}$, the constant terms in the equation must sum to zero and the terms multiplying each element of $\mathbf{x}$ must also sum to zero. These conditions imply

$$\frac{\partial A(\tau)}{\partial\tau} = \alpha + (\boldsymbol{\kappa}\overline{\mathbf{x}})'\,\mathbf{B}(\tau) - \tfrac{1}{2}\sum_{i=1}^{n}[\boldsymbol{\Sigma}'\mathbf{B}(\tau)]_i^2\, s_{0i} \tag{17.18}$$

$$\frac{\partial \mathbf{B}(\tau)}{\partial\tau} = \boldsymbol{\beta} - \boldsymbol{\kappa}'\mathbf{B}(\tau) - \tfrac{1}{2}\sum_{i=1}^{n}[\boldsymbol{\Sigma}'\mathbf{B}(\tau)]_i^2\, \mathbf{s}_{1i} \tag{17.19}$$

where $[\boldsymbol{\Sigma}'\mathbf{B}(\tau)]_i$ is the i^{th} element of the $n \times 1$ vector $\boldsymbol{\Sigma}'\mathbf{B}(\tau)$. Equations (17.18) and (17.19) are a system of first-order ordinary differential equations that can be solved subject to the boundary conditions $A(0) = 0$ and $\mathbf{B}(0) = \mathbf{0}$. In some cases, such as a multiple state variable version of the Vasicek model (where $\mathbf{s}_{1i} = \mathbf{0}\ \forall i$), there exist closed-form solutions.[6] In other cases, fast and accurate numerical solutions to these ordinary differential equations can be computed using techniques such as a Runge-Kutta algorithm.

While affine term structure models require that the state variables' risk-neutral expected changes be affine in the state variables, there is more flexibility regarding the state variables' drifts under the physical measure. Note that the state variables' expected change under the physical measure is

$$\mathbf{a}(t,\mathbf{x}) = \boldsymbol{\kappa}(\overline{\mathbf{x}}-\mathbf{x}) + \boldsymbol{\Sigma}\sqrt{\mathbf{s}(\mathbf{x})}\boldsymbol{\Theta} \tag{17.20}$$

[6] Examples include (Langeteig 1980), (Pennacchi 1991), and (Jegadeesh and Pennacchi 1996).

so that specification of the market prices of risk, $\mathbf{\Theta}$, is required to determine the physical drifts of the state variables. Qiang Dai and Kenneth Singleton (Dai and Singleton 2000) study the "completely affine" case where both the physical and risk-neutral drifts are affine, while Gregory Duffee (Duffee 2002) and Jefferson Duarte (Duarte 2004) consider extensions of the physical drifts that permit nonlinearities.[7] Because the means, volatilities, and risk premia of bond prices estimated from time series data depend on the physical moments of the state variables, the flexibility in choosing the parametric form for $\mathbf{\Theta}$ can allow the model to better fit historical bond price data.

Example: Independent Factors

Consider the special case where $\boldsymbol{\kappa}$ and $\mathbf{\Sigma}$ are $n \times n$ diagonal matrices and the $n \times 1$ vector $\mathbf{s}_{1i}$ has all of its elements equal to zero except for its i^{th} element. These assumptions imply that the risk-neutral drift term of each state variable depends only on its own level and that the state variables' covariance matrix, $\mathbf{b}(t, \mathbf{x})\,\mathbf{b}(t, \mathbf{x})' = \mathbf{\Sigma}\mathbf{s}(\mathbf{x})\,\mathbf{\Sigma}'$, is diagonal. Thus, this case is one where the processes for the state variables are independent of each other. Further, for simplicity, let $r(t, \mathbf{x}) = \alpha + \boldsymbol{\beta}'\mathbf{x} = \mathbf{e}'\mathbf{x}$, so that $\alpha = 0$ and $\boldsymbol{\beta} = \mathbf{e}$ is an $n \times 1$ vector of ones.[8] Given these parametric restrictions, the interest rate is the sum of independent state variables and the bond valuation equation (17.7) becomes

[7] Dai and Singleton analyze $\mathbf{\Theta} = \sqrt{\mathbf{s}(\mathbf{x})}\boldsymbol{\lambda}_1$ where $\boldsymbol{\lambda}_1$ is an $n \times 1$ vector of constants. Duffee considers the "essentially affine" modeling of the market price of risk of the form $\mathbf{\Theta} = \sqrt{\mathbf{s}(\mathbf{x})}\boldsymbol{\lambda}_1 + \sqrt{\mathbf{s}(\mathbf{x})^{-}}\boldsymbol{\lambda}_2\mathbf{x}$, where $\mathbf{s}(\mathbf{x})^{-}$ is an $n \times n$ diagonal matrix whose i^{th} element equals $\left(s_{oi} + \mathbf{s}_{1i}'\mathbf{x}\right)^{-1}$ if inf $\left(s_{oi} + \mathbf{s}_{1i}'\mathbf{x}\right) > 0$ and zero otherwise, and $\boldsymbol{\lambda}_2$ is an $n \times n$ matrix of constants. This specification allows time variation in the market prices of risk for Gaussian state variables (such as state variables that follow Ornstein-Uhlenbeck processes), allowing their signs to switch over time. Duarte extends Duffee's modeling to add a square root term. This "semiaffine square root" model takes the form $\mathbf{\Theta} = \mathbf{\Sigma}^{-1}\boldsymbol{\lambda}_0 + \sqrt{\mathbf{s}(\mathbf{x})}\boldsymbol{\lambda}_1 + \sqrt{\mathbf{s}(\mathbf{x})^{-}\boldsymbol{\lambda}_2\mathbf{x}}$ where $\boldsymbol{\lambda}_0$ is an $n \times 1$ vector of constants. See also work by Patrick Cheridito, Damir Filipovic, and Robert Kimmel (Cheridito, Filipovic, and Kimmel 2003) for extensions in modeling the market price of risk for affine models.

[8] The assumptions regarding α and $\boldsymbol{\beta}$ are not restrictive to the results derived below. A nonzero α would add a multiplicative constant to bond prices and each state variable can be normalized by its $\boldsymbol{\beta}$ element to give a similar result.

$$
\begin{aligned}
P(t,T,\mathbf{x}) &= \widehat{E}_t\left[e^{-\int_t^T r(s,\mathbf{x})ds} \times 1\right] \\
&= \widehat{E}_t\left[e^{-\int_t^T \mathbf{e}'\mathbf{x}ds}\right] \\
&= \prod_{i=1}^{n} \widehat{E}_t\left[e^{-\int_t^T x_i(s)ds}\right]
\end{aligned}
\tag{17.21}
$$

where the last line in (17.21) results from the independence assumption. The insight from (17.21) is that this multifactor term structure model can be interpreted as the product of n single-factor term structure models, where each state variable, x_i, is analogous to a different interest rate. For example, if $\mathbf{s}_i(\mathbf{x}) = s_{oi}$, so that x_i follows an Ornstein-Uhlenbeck process, then $\widehat{E}_t\left[\exp\left(-\int_t^T x_i(s)\,ds\right)\right] = \exp\left[A_i(\tau) + B_i(\tau)\,x_i\right]$ where the functions $A_i(\tau)$ and $B_i(\tau)$ solve simplified versions of (17.18) and (17.19) and take similar forms to the Vasicek model formula in (9.41).[9] Another state variable, say, x_j, could have $\mathbf{s}_j(\mathbf{x}) = s_{1j}x_j$, so that it follows a square root constant elasticity of variance process. For this state variable, $\widehat{E}_t\left[\exp\left(-\int_t^T x_j(s)\,ds\right)\right] = \exp\left[A_j(\tau) + B_j(\tau)\,x_j\right]$ where the functions $A_j(\tau)$ and $B_j(\tau)$ satisfy simple versions of (17.18) and (17.19) and have solutions similar to the CIR model formula in (13.51).[10] Thus, using these prior single-factor model results, (17.21) can be written as

$$
P(t,T,\mathbf{x}) = \prod_{i=1}^{n} \exp\left[A_i(\tau) + B_i(\tau)\,x_i\right] \tag{17.22}
$$

Whether the assumption that state variables are independent is reasonable depends on the particular empirical context in which a term structure model is being used. Typically, there is a trade-off between more general correlation structures and model simplicity. Gaussian state variables (e.g., those following

[9] Due to slightly different notation, $A_i(\tau)$ equals $\ln[A(\tau)]$ in (9.43) and $B_i(\tau)$ equals $-B(\tau)$ in (9.42).

[10] Because of slightly different notation, $A_j(\tau)$ equals $\ln[A(\tau)]$ in (13.52) and $B_j(\tau)$ equals $-B(\tau)$ in (13.53).

an Ornstein-Uhlenbeck process) allow for general correlation structures but do not restrict the state variables from becoming negative. State variables following square root processes can be restricted to maintain positive values but may be incapable of displaying negative correlation.

17.1.2 Quadratic Gaussian Models

Another class of models assumes that the yields of zero-coupon bonds are quadratic functions of normally distributed (Gaussian) state variables. Markus Leippold and Liuren Wu (Leippold and Wu 2002) provide a detailed discussion of these models. We can express the assumption that yields are a quadratic function of state variables by stating

$$Y(t,T,\mathbf{x})\,\tau = A(\tau) + \mathbf{B}(\tau)'\,\mathbf{x} + \mathbf{x}'\mathbf{C}(\tau)\,\mathbf{x} \tag{17.23}$$

where $\mathbf{C}(\tau)$ is an $n\times n$ matrix and, with no loss of generality, can be assumed to be symmetric. Similar to our analysis of affine models, since $P(T,T,\mathbf{x}) = 1$, we must have $A(0) = 0$, $\mathbf{B}(0)$ equal to an $n\times 1$ vector of zeros, and $\mathbf{C}(0)$ equal to an $n\times n$ matrix of zeros. In addition, the yield on a bond of instantaneous maturity must be of the form $r(t,\mathbf{x}) = \alpha + \boldsymbol{\beta}'\mathbf{x} + \mathbf{x}'\boldsymbol{\gamma}\mathbf{x}$, where $\alpha = \partial A(0)/\partial\tau$, $\boldsymbol{\beta} = \partial\mathbf{B}(0)/\partial\tau$, and $\boldsymbol{\gamma} = \partial\mathbf{C}(0)/\partial\tau$ is an $n\times n$ symmetric matrix of constants. Note that if $\boldsymbol{\gamma}$ is a positive semidefinite matrix and $\alpha - \frac{1}{4}\boldsymbol{\beta}'\boldsymbol{\gamma}^{-1}\boldsymbol{\beta} \geq 0$, then the interest rate can be restricted from becoming negative.[11] Substituting $P(t,T,\mathbf{x}) = \exp\left(-A(\tau) - \mathbf{B}(\tau)'\,\mathbf{x} - \mathbf{x}'\mathbf{C}(\tau)\,\mathbf{x}\right)$ into the general partial differential equation (17.6), we obtain

[11] The lower bound for $r(t)$ is $\alpha - \frac{1}{4}\boldsymbol{\beta}'\boldsymbol{\gamma}^{-1}\boldsymbol{\beta}$, which occurs when $\mathbf{x} = -\frac{1}{2}\boldsymbol{\gamma}^{-1}\boldsymbol{\beta}$.

$$\begin{aligned}
&\tfrac{1}{2}\left[\left[\mathbf{B}(\tau)+2\mathbf{C}(\tau)\mathbf{x}\right]'\mathbf{b}(t,\mathbf{x})\mathbf{b}(t,\mathbf{x})'\left[\mathbf{B}(\tau)+2\mathbf{C}(\tau)\mathbf{x}\right]\right] \\
&-\mathrm{Trace}\left[\mathbf{b}(t,\mathbf{x})'\mathbf{C}(\tau)\mathbf{b}(t,\mathbf{x})\right] \\
&-\left[\mathbf{a}(t,\mathbf{x})-\mathbf{b}(t,\mathbf{x})\boldsymbol{\Theta}\right]'\left[\mathbf{B}(\tau)+2\mathbf{C}(\tau)\mathbf{x}\right] \\
&+\frac{\partial A(\tau)}{\partial\tau}+\frac{\partial\mathbf{B}(\tau)'}{\partial\tau}\mathbf{x}+\mathbf{x}'\frac{\partial\mathbf{C}(\tau)}{\partial\tau}\mathbf{x} \\
=\;&\alpha+\boldsymbol{\beta}'\mathbf{x}+\mathbf{x}'\boldsymbol{\gamma}\mathbf{x} \qquad (17.24)
\end{aligned}$$

In addition to yields being quadratic in the state variables, quadratic Gaussian models then assume that the vector of state variables, $\mathbf{x}$, has a multivariate normal (Gaussian) distribution. Specifically, it is assumed that $\mathbf{x}$ follows a multivariate Ornstein-Uhlenbeck process:

$$\mathbf{a}(t,\mathbf{x})-\mathbf{b}(t,\mathbf{x})\boldsymbol{\Theta}=\boldsymbol{\kappa}(\overline{\mathbf{x}}-\mathbf{x}) \qquad (17.25)$$

$$\mathbf{b}(t,\mathbf{x})=\boldsymbol{\Sigma} \qquad (17.26)$$

Substituting these assumptions into the partial differential equation (17.24), one obtains

$$\begin{aligned}
&\tfrac{1}{2}\left[\left[\mathbf{B}(\tau)+2\mathbf{C}(\tau)\mathbf{x}\right]'\boldsymbol{\Sigma}\boldsymbol{\Sigma}'\left[\mathbf{B}(\tau)+2\mathbf{C}(\tau)\mathbf{x}\right]\right] \\
&-\mathrm{Trace}\left[\boldsymbol{\Sigma}'\mathbf{C}(\tau)\boldsymbol{\Sigma}\right]-\left[\boldsymbol{\kappa}(\overline{\mathbf{x}}-\mathbf{x})\right]'\left[\mathbf{B}(\tau)+2\mathbf{C}(\tau)\mathbf{x}\right] \\
&+\frac{\partial A(\tau)}{\partial\tau}+\frac{\partial\mathbf{B}(\tau)'}{\partial\tau}\mathbf{x}+\mathbf{x}'\frac{\partial\mathbf{C}(\tau)}{\partial\tau}\mathbf{x} \\
=\;&\alpha+\boldsymbol{\beta}'\mathbf{x}+\mathbf{x}'\boldsymbol{\gamma}\mathbf{x} \qquad (17.27)
\end{aligned}$$

For this equation to hold for all values of $\mathbf{x}$, it must be the case that the sums of the equation's constant terms, the terms proportional to the elements of $\mathbf{x}$, and the terms that are products of the elements of $\mathbf{x}$ must each equal zero. This leads to the system of first-order ordinary differential equations

$$\frac{\partial A(\tau)}{\partial \tau} = \alpha + (\boldsymbol{\kappa}\overline{\mathbf{x}})' \mathbf{B}(\tau) - \tfrac{1}{2}\mathbf{B}(\tau)' \boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{B}(\tau) + \text{Trace}\left[\boldsymbol{\Sigma}'\mathbf{C}(\tau)\boldsymbol{\Sigma}\right] \tag{17.28}$$

$$\frac{\partial \mathbf{B}(\tau)}{\partial \tau} = \boldsymbol{\beta} - \boldsymbol{\kappa}'\mathbf{B}(\tau) - 2\mathbf{C}(\tau)' \boldsymbol{\Sigma}'\boldsymbol{\Sigma}\mathbf{B}(\tau) + 2\mathbf{C}(\tau)' \boldsymbol{\kappa}\overline{\mathbf{x}} \tag{17.29}$$

$$\frac{\partial \mathbf{C}(\tau)}{\partial \tau} = \boldsymbol{\gamma} - 2\boldsymbol{\kappa}'\mathbf{C}(\tau) - 2\mathbf{C}(\tau)' \boldsymbol{\Sigma}\boldsymbol{\Sigma}'\mathbf{C}(\tau) \tag{17.30}$$

which are solved subject to the aforementioned boundary conditions, $A(0) = 0$, $\mathbf{B}(0) = \mathbf{0}$, and $\mathbf{C}(0) = \mathbf{0}$.

Dong-Hyun Ahn, Robert Dittmar, and Ronald Gallant (Ahn, Dittmar, and Gallant 2002) show that the models of Francis Longstaff (Longstaff 1989), David Beaglehole and Mark Tenney (Beaglehole and Tenney 1992), and George Constantinides (Constantinides 1992) are special cases of quadratic Gaussian models. They also demonstrate that since quadratic Gaussian models allow a nonlinear relationship between yields and state variables, these models can outperform affine models in explaining historical bond yield data.

However, quadratic Gaussian models are more difficult to estimate from historical data because, unlike affine models, there is not a one-to-one mapping between bond yields and the elements of the vector of state variables. For example, suppose that at a given point in time, we observed bond yields of n different maturities, say, $Y(t, T_i, \mathbf{x})$, $i = 1, ..., n$. Denoting $\tau_i = T_i - t$, if yields are affine functions of the state variables, then $Y(t, T_i, \mathbf{x})\tau_i = A(\tau_i) + \mathbf{B}(\tau_i)' \mathbf{x}$, $i = 1, .., n$, represents a set of n linear equations in the n elements of the state variable $\mathbf{x}$. Solving these equations for the state variables x_1, x_2, ..., x_n effectively allows one to observe the individual state variables from the observed yields. By observing a time series of these state variables, the parameters of their physical process could be estimated.

This approach cannot be used when yields are quadratic functions of the state variables since with $Y(t, T_i, \mathbf{x})\tau_i = A(\tau_i) + \mathbf{B}(\tau_i)'\mathbf{x} + \mathbf{x}'\mathbf{C}(\tau_i)\mathbf{x}$, there is not a one-to-one mapping between yields and state variables x_1, x_2, ..., x_n. There are multiple values of the state variable vector, $\mathbf{x}$, consistent with the set of yields.[12] This difficulty requires a different approach to inferring the most likely state variable vector. Ahn, Dittmar, and Gallant use an efficient method of moments technique that simulates the state variable, $\mathbf{x}$, to estimate the state variable vector that best fits the data.

17.1.3 Other Equilibrium Models

Term structure models have been modified to allow state variable processes to differ from strict diffusions. Such models can no longer rely on the Black-Scholes hedging argument to identify market prices of risk and a risk-neutral pricing measure. Because fixed-income markets may not be dynamically complete, these models need to make additional assumptions regarding the market prices of risks that cannot be hedged.

A number of researchers, including Chang-Mo Ahn and Howard Thompson (Ahn and Thompson 1988), Sanjiv Das and Silverio Foresi (Das and Foresi 1996), Darrell Duffie, Jun Pan, and Kenneth Singleton (Duffie, Pan, and Singleton 2000), Sanjiv Das (Das 2002), and George Chacko and Sanjiv Das (Chacko and Das 2002), have extended equilibrium models to allow state variables to follow jump-diffusion processes. An interesting application of a model with jumps in a short-term interest rate is presented by Monika Piazzesi (Piazzesi 2005b) who studies the Federal Reserve's changes in the target federal funds rate.

Other affine equilibrium models have been set in discrete time, where the assumed existence of a discrete-time pricing kernel allows one to find solutions

[12] For example, if $n = 1$, there are two state variable roots of the quadratic yield equation.

for equilibrium bond prices that have a recursive structure. Examples of models of this type include work by Tong-Sheng Sun (Sun 1992), David Backus and Stanley Zin (Backus and Zin 1994), V. Cvsa and Peter Ritchken (Cvsa and Ritchken 2001), and Qiang Dai, Anh Le, and Kenneth Singleton (Dai, Le, and Singleton 2006). Term structure models also have been generalized to include discrete regime shifts in the processes followed by state variables. See work by Vasant Naik and Moon Hoe Lee (Naik and Lee 1997) and Ravi Bansal and Hao Zhou (Bansal and Zhou 2002) for models of this type.

Let us now turn to fixed-income models whose primary purpose is not to determine the term structure of zero-coupon bond prices as a function of state variables. Rather, their objective is to determine the value of bond and interest rate-related derivatives as a function of a given term structure of bond prices.

17.2 Valuation Models for Interest Rate Derivatives

Models for valuing bonds and bond derivatives have different uses. The equilibrium models of the previous section can provide insights as to the nature of term structure movements. They allow us to predict how factor dynamics influence the prices of bonds of different maturities. Equilibrium models may also be of practical use to bond traders who wish to identify bonds of particular maturities that appear to be over- or underpriced based on their predicted model valuations. Such information could suggest profitable bond trading strategies.

However, bond prices are modeled for other objectives, such as the pricing of derivatives whose payoffs depend on the future prices of bonds or yields. Equilibrium models may be less than satisfactory for this purpose because it is bond derivatives, not the underlying bond prices themselves, that one wishes to value. In this context, one would like to use the observed market prices for bonds as an input into the valuation formulas for derivatives, not model the value of the underlying bonds themselves. For such a derivative-pricing exercise, one would like the model to "fit," or be consistent with, the observed market prices of the underlying bonds. The models that we will now consider are designed to have this feature.

17.2.1 Heath-Jarrow-Morton Models

The approach by David Heath, Robert Jarrow, and Andrew Morton (Heath, Jarrow, and Morton 1992), hereafter referred to as HJM, differs from the previous equilibrium term structure models because it does not begin by specifying a set of state variables, $\mathbf{x}$, that determines the current term structure of bond prices. Rather, their approach takes the initial term structure of bond prices as given (observed) and then specifies how this term structure evolves in the future in order to value derivatives whose payoffs depend on future term structures. Because models of this type do not derive the term structure from more basic state variables, they cannot provide insights regarding how economic fundamentals determine the maturity structure of zero-coupon bond prices. Instead, HJM models are used to value fixed-income derivative securities: securities such as bond and interest rate options whose payoffs depend on future bond prices or yields.

An analogy to the HJM approach can be drawn from the risk-neutral valuation of equity options. Recall that in Chapter 10, equation (10.50), we assumed that the risk-neutral process for the price of a stock, $S(t)$, followed geometric Brownian motion, making this price lognormally distributed under the risk-neutral measure. From this assumption, and given the initial price of the stock, $S(t)$, the Black-Scholes formula for the value of a call option written on this stock was derived in equations (10.54) and (10.55). Note that we did not attempt to determine the initial value of the stock in terms of some fundamental state variables, say $S(t, \mathbf{x})$. Rather, the initial stock price, $S(t)$, was taken as given and an assumption about this stock price's volatility, namely, that it was constant over time, was made.

The HJM approach to valuing fixed-income derivatives is similar but slightly more complex because it takes as given the entire initial term structure of bond prices, $P(t,T)\ \forall T \geq t$, not just a single asset (stock) price. It then assumes risk-neutral processes for how the initial observed bond prices change over time and does not attempt to derive these initial prices in terms of state variables, say, $P(t,T,\mathbf{x})$. However, the way that HJM specify the processes followed by bond prices is somewhat indirect. They begin by specifying processes for bond *forward rates*. A fundamental result of the HJM analysis is to show that, in the absence of arbitrage, there must be a particular relationship between the drift and volatility parameters of forward rate processes and that only an assumption regarding the form of forward rate volatilities is needed for pricing derivatives.

Let us start by defining forward rates. Recall from Chapter 7 that a forward contract is an agreement between two parties where the long (*short*) party agrees to purchase (*deliver*) an underlying asset in return for paying (*receiving*) the forward price. Consider a forward contract agreed to at date t, where the contract matures at date $T \geq t$ and the underlying asset is a zero-coupon bond that matures at date $T+\tau$ where $\tau \geq 0$. Let $F(t,T,\tau)$ be the equilibrium forward price agreed to by the parties. Then this contract requires the long party to pay $F(t,T,\tau)$ at date T in return for receiving a cashflow of \$1 (the zero-coupon bond's maturity value) at date $T+\tau$. In the absence of arbitrage, the value of these two cashflows at date t must sum to zero, implying

$$-F(t,T,\tau)\,P(t,T)+P(t,T+\tau)=0 \tag{17.31}$$

so that the equilibrium forward price equals the ratio of the bond prices maturing at dates $T+\tau$ and T, $F(t,T,\tau) = P(t,T+\tau)/P(t,T)$. From this forward price a continuously compounded forward rate, $f(t,T,\tau)$, is defined as

$$e^{-f(t,T,\tau)\tau} \equiv F(t,T,\tau) = \frac{P(t,T+\tau)}{P(t,T)} \tag{17.32}$$

$f(t,T,\tau) = -(\ln[P(t,T+\tau)/P(t,T)])/\tau$ is the implicit per-period rate of return (interest rate) that the long party earns by investing $\$F(t,T,T+\tau)$ at date T and by receiving \$1 at date $T+\tau$. Now consider the case of such a forward contract where the underlying bond matures very shortly (e.g., the next day or instant) after the maturity of the forward contract. This permits us to define an instantaneous forward rate as

$$f(t,T) \equiv \lim_{\tau\to 0} f(t,T,\tau) = \lim_{\tau\to 0} -\frac{\ln[P(t,T+\tau)] - \ln[P(t,T)]}{\tau} = -\frac{\partial \ln[P(t,T)]}{\partial T} \tag{17.33}$$

Equation (17.33) is a simple differential equation that can be solved to obtain

$$P(t,T) = e^{-\int_t^T f(t,s)ds} \tag{17.34}$$

Since this bond's continuously compounded yield to maturity is defined from the relation $P(t,T) = e^{-Y(t,T)(T-t)}$, we can write $Y(t,T) = \frac{1}{T-t}\int_t^T f(t,s)\,ds$. Thus, a bond's yield equals the average of the instantaneous forward rates for horizons out to the bond's maturity. In particular, the yield on an instantaneous-maturity bond is given by $r(t) = f(t,t)$.

Because the term structure of instantaneous forward rates, $f(t,T)\,\forall T \geq t$, can be determined from the term structure of bond prices, $P(t,T)\,\forall T \geq t$, or yields, $Y(t,T)\,\forall T \geq t$, specifying the evolution of forward rates over time is equivalent to specifying the dynamics of bond prices. HJM assume that forward rates for all horizons are driven by a finite-dimensional Brownian motion:

$$df(t,T) = \alpha(t,T)\,dt + \boldsymbol{\sigma}(t,T)'\,\mathbf{dz} \tag{17.35}$$

where $\boldsymbol{\sigma}(t,T)$ is an $n \times 1$ vector of volatility functions and $\mathbf{dz}$ is an $n \times 1$ vector of independent Brownian motions. Note that since there are an infinite number of instantaneous forward rates, one for each future horizon, equation (17.35) represents infinitely many processes that are driven by the same n Brownian motions.

Importantly, the absence of arbitrage places restrictions on $\alpha(t,T)$ and $\boldsymbol{\sigma}(t,T)$. To show this, let us start by deriving the process followed by bond prices, $P(t,T)$, implied by the forward rate processes. Note that since $\ln[P(t,T)] = -\int_t^T f(t,s)\,ds$, if we differentiate with respect to date t, we find that the process followed by the log bond price is

$$\begin{aligned} d\ln[P(t,T)] &= f(t,t)\,dt - \int_t^T df(t,s)\,ds \\ &= r(t)\,dt - \int_t^T \left[\alpha(t,s)\,dt + \boldsymbol{\sigma}(t,s)'\,\mathbf{dz}(t)\right] ds \end{aligned} \tag{17.36}$$

Fubini's theorem allows us to switch the order of integration:

$$\begin{aligned} d\ln[P(t,T)] &= r(t)\,dt - \int_t^T \alpha(t,s)\,ds dt - \int_t^T \boldsymbol{\sigma}(t,s)'\,ds\mathbf{dz}(t) \\ &= r(t)\,dt - \alpha_I(t,T)\,dt - \boldsymbol{\sigma}_I(t,T)'\,\mathbf{dz}(t) \end{aligned} \tag{17.37}$$

where we have used the shorthand notation $\alpha_I(t,T) \equiv \int_t^T \alpha(t,s)\,ds$ and $\boldsymbol{\sigma}_I(t,T) \equiv \int_t^T \boldsymbol{\sigma}(t,s)\,ds$ to designate these integrals that are known functions as of date t. Using Itô's lemma we can derive the bond's rate of return process from the log process in (17.37):

$$\frac{dP(t,T)}{P(t,T)} = \left[r(t) - \alpha_I(t,T) + \frac{1}{2}\boldsymbol{\sigma}_I(t,T)'\boldsymbol{\sigma}_I(t,T)\right]dt - \boldsymbol{\sigma}_I(t,T)'\mathbf{dz} \quad (17.38)$$

Now recall from (17.5) that the absence of arbitrage requires that the bond's expected rate of return equal the instantaneous risk-free return plus the product of the bond's volatilities and the market prices of risk. This is written as

$$r(t) - \alpha_I(t,T) + \frac{1}{2}\boldsymbol{\sigma}_I(t,T)'\boldsymbol{\sigma}_I(t,T) = r(t) - \boldsymbol{\Theta}(t)'\boldsymbol{\sigma}_I(t,T) \quad (17.39)$$

or

$$\alpha_I(t,T) = \frac{1}{2}\boldsymbol{\sigma}_I(t,T)'\boldsymbol{\sigma}_I(t,T) + \boldsymbol{\Theta}(t)'\boldsymbol{\sigma}_I(t,T) \quad (17.40)$$

Equations (17.38) and (17.40) show that the bond price process depends only on the instantaneous risk-free rate, the volatilities of the forward rates, and the market prices of risk. This no-arbitrage condition also has implications for the risk-neutral process followed by forward rates. If we substitute $\mathbf{dz} = \mathbf{d}\widehat{\mathbf{z}} - \boldsymbol{\Theta}(t)\,dt$ in (17.35), we obtain

$$\begin{aligned} df(t,T) &= \left[\alpha(t,T) - \boldsymbol{\sigma}(t,T)'\boldsymbol{\Theta}(t)\right]dt + \boldsymbol{\sigma}(t,T)'\mathbf{d}\widehat{\mathbf{z}} \\ &= \widehat{\alpha}(t,T)\,dt + \boldsymbol{\sigma}(t,T)'\mathbf{d}\widehat{\mathbf{z}} \end{aligned} \quad (17.41)$$

where $\widehat{\alpha}(t,T) \equiv \alpha(t,T) - \boldsymbol{\sigma}(t,T)'\boldsymbol{\Theta}(t)$ is the risk-neutral drift observed at date t for the forward rate at date T. Define $\widehat{\alpha}_I(t,T) \equiv \int_t^T \widehat{\alpha}(t,s)\,ds$ as the integral over the drifts across all forward rates from date t to date T. Then using (17.40) we have

$$\begin{aligned}
\widehat{\alpha}_I(t,T) &= \int_t^T \widehat{\alpha}(t,s)\,ds = \int_t^T \alpha(t,s)\,ds - \int_t^T \boldsymbol{\sigma}(t,s)'\,ds\boldsymbol{\Theta}(t) \\
&= \alpha_I(t,T) - \boldsymbol{\Theta}(t)'\boldsymbol{\sigma}_I(t,T) \\
&= \frac{1}{2}\boldsymbol{\sigma}_I(t,T)'\,\boldsymbol{\sigma}_I(t,T) + \boldsymbol{\Theta}(t)'\boldsymbol{\sigma}_I(t,T) - \boldsymbol{\Theta}(t)'\boldsymbol{\sigma}_I(t,T) \\
&= \frac{1}{2}\boldsymbol{\sigma}_I(t,T)'\,\boldsymbol{\sigma}_I(t,T) \qquad (17.42)
\end{aligned}$$

or $\int_t^T \widehat{\alpha}(t,s)\,ds = \frac{1}{2}\left(\int_t^T \boldsymbol{\sigma}(t,s)\,ds\right)'\left(\int_t^T \boldsymbol{\sigma}(t,s)\,ds\right)$. This shows that in the absence of arbitrage, the risk-neutral drifts of forward rates are completely determined by their volatilities. Indeed, if we differentiate $\widehat{\alpha}_I(t,T)$ with respect to T to recover $\widehat{\alpha}(t,T)$, we obtain

$$\begin{aligned}
df(t,T) &= \boldsymbol{\sigma}(t,T)'\,\boldsymbol{\sigma}_I(t,T)\,dt + \boldsymbol{\sigma}(t,T)'\,\mathbf{d}\widehat{\mathbf{z}} \qquad (17.43) \\
&= \left(\boldsymbol{\sigma}(t,T)'\int_t^T \boldsymbol{\sigma}(t,s)\,ds\right)dt + \boldsymbol{\sigma}(t,T)'\,\mathbf{d}\widehat{\mathbf{z}}
\end{aligned}$$

Equation (17.43) has an important implication, namely, that if we want to model the risk-neutral dynamics of forward rates in order to price fixed-income derivatives, we need only specify the form of the forward rates' volatility functions.[13] One can also use (17.43) to derive the risk-neutral dynamics of the instantaneous-maturity interest rate, $r(t) = f(t,t)$, which is required for discounting risk-neutral payoffs. Suppose dates are ordered such that $0 \leq t \leq T$. In integrated form, (17.43) becomes

$$f(t,T) = f(0,T) + \int_0^t \boldsymbol{\sigma}(u,T)'\,\boldsymbol{\sigma}_I(u,T)\,du + \int_0^t \boldsymbol{\sigma}(u,T)'\,\mathbf{d}\widehat{\mathbf{z}}(u) \qquad (17.44)$$

[13] In general, these volatility functions may be stochastic, as they could be specified to depend on current levels of the forward rates, that is, $\boldsymbol{\sigma}(t,T,f(t,T))$.

and for $r(t) = f(t,t)$, this becomes

$$r(t) = f(0,t) + \int_0^t \boldsymbol{\sigma}(u,t)' \boldsymbol{\sigma}_I(u,t)\, du + \int_0^t \boldsymbol{\sigma}(u,t)' \mathbf{d}\widehat{\mathbf{z}}(u) \tag{17.45}$$

Differentiating with respect to t leads to[14]

$$\begin{aligned} dr(t) &= \frac{\partial f(0,t)}{\partial t} dt + \boldsymbol{\sigma}(t,t)' \boldsymbol{\sigma}_I(t,t)\, dt + \int_0^t \frac{\partial \boldsymbol{\sigma}(u,t)' \boldsymbol{\sigma}_I(u,t)}{\partial t} du dt \\ &\quad + \int_0^t \frac{\partial \boldsymbol{\sigma}(u,t)'}{\partial t} \mathbf{d}\widehat{\mathbf{z}}(u)\, dt + \boldsymbol{\sigma}(t,t)' \mathbf{d}\widehat{\mathbf{z}} \\ &= \frac{\partial f(0,t)}{\partial t} dt + \int_0^t \left[\boldsymbol{\sigma}(u,t)' \boldsymbol{\sigma}(u,t) + \frac{\partial \boldsymbol{\sigma}(u,t)'}{\partial t} \boldsymbol{\sigma}_I(u,t) \right] du dt \\ &\quad + \int_0^t \frac{\partial \boldsymbol{\sigma}(u,t)'}{\partial t} \mathbf{d}\widehat{\mathbf{z}}(u)\, dt + \boldsymbol{\sigma}(t,t)' \mathbf{d}\widehat{\mathbf{z}} \end{aligned} \tag{17.46}$$

where we have used the fact that $\boldsymbol{\sigma}_I(t,t) = \mathbf{0}$ and $\partial \boldsymbol{\sigma}_I(u,t)/\partial t = \boldsymbol{\sigma}(u,t)$.

With these results, one can now value fixed-income derivatives. As an example, define $C(t)$ as the current date t price of a European-type contingent claim that has a payoff at date T. This payoff is assumed to depend on the forward rate curve (equivalently, the term structure of bond prices or yields) at date T, which we write as $C(T, f(T, T+\delta))$ where $\delta \geq 0$. The contingent claim's risk-neutral valuation equation is

$$C(t, f(t, t+\delta)) = \widehat{E}_t \left[e^{-\int_t^T r(s)ds} C(T, f(T, T+\delta)) \mid f(t, t+\delta), \forall \delta \geq 0 \right] \tag{17.47}$$

where the expectation is conditioned on information of the current date t forward rate curve, $f(t, t+\delta)$ $\forall \delta \geq 0$. Equation (17.47) is the risk-neutral expectation of the claim's discounted payoff, conditional on information of all currently observed forward rates. In this manner, the contingent claim's formula can be

[14] Note that the dynamics of dr are more complicated than simply setting $T = t$ in equation (17.43), because both arguments of $f(t,t) = r(t)$ are varying simultaneously. Equation (17.46) is equivalent to $dr = df(t,t) + \frac{\partial f(t,u)}{\partial u} |_{u \to t}\, dt$.

assured of fitting the current term structure of interest rates, since the forward rate curve, $f(t,t+\delta)$, is an input. Only for special cases regarding the type of contingent claim and the assumed forward rate volatilities can the expectation in (17.47) be computed analytically. In general, it can be computed by a Monte Carlo simulation of a discrete-time analog to the continuous-time, risk-neutral forward rate and instantaneous interest rate processes in (17.43) and (17.46).[15]

Valuing American-type contingent claims using the HJM approach can be more complicated because, in general, one needs to discretize forward rates to produce a lattice (e.g., binomial tree) and check the nodes of the lattice to see if early exercise is optimal.[16] However, HJM forward rates will not necessarily follow Markov processes. From (17.43) and (17.46), one can see that if the forward rate volatility functions are specified to depend on the level of forward rates themselves, $\sigma(t,s,f(t,s))$, or the instantaneous risk-free rate, $\sigma(t,s,r(t))$, then the evolution of $f(t,T)$ and $r(t)$ depends on the entire history of forward rates between two dates such as 0 and t. It will be impossible to express forward rates as $f(0,T,\mathbf{x}(0))$ and $f(t,T,\mathbf{x}(t))$ where $\mathbf{x}(t)$ is a set of finite state variables.[17] Non-Markov processes lead to lattice structures where the nodes do not recombine. This can make computation extremely time consuming because the number of nodes grows exponentially (rather than linearly in the case of recombining nodes) with the number of time steps. Hence, to value American contingent claims using the HJM framework, it is highly desirable to pick volatility structures that lead to forward rate processes that are Markov.[18]

[15] An example is presented by Kaushik Amin and Andrew Morton (Amin and Morton 1994). They value Eurodollar futures and options assuming different one-factor ($n=1$) specifications for forward rate volatilities. Their models are nested in the functional form $\sigma(t,T)=[\sigma_0+\sigma_1(T-t)]\,e^{-\alpha(T-t)}f(t,T)^{\gamma}$.

[16] Recall that this method was used in Chapter 7 to value an American option.

[17] The reason why one may want to assume that forward rate volatilities depend on their own level is to preclude negative forward rates, a necessary condition if currency is not to dominate bonds in a nominal term structure model. For example, similar to the square root model of Cox, Ingersoll, and Ross, one could specify $\sigma(t,T)=\overline{\sigma}(t,T)\,f(t,T)^{\frac{1}{2}}$ or $\sigma(t,T)=\overline{\sigma}(t,T)\,r(t)^{\frac{1}{2}}$ where $\overline{\sigma}(t,T)$ is a deterministic function.

[18] Note, also, that non-Markov short rate and forward rate processes imply that contingent

The next section gives two examples of HJM models that are Markov in a finite number of state variables.

Examples: Markov HJM Models

General conditions on forward rate volatilities that lead to Markov structures are discussed in Koji Inui and Masaaki Kijima (Inui and Kijima 1998). In this section we give two different examples of Markov HJM models. The first is an example where forward rates, including the instantaneous-maturity interest rate, are Markov in one state variable. In the second example, rates are Markov in two state variables. In both examples, it is assumed that $n = 1$, so that there is a single Brownian motion process driving all forward rates.

Our first example assumes forward rate volatilities are deterministic. As shown by Andrew Carverhill (Carverhill 1994), this assumption results in HJM models that are Markov in one state variable. Here we consider a particular case of deterministic forward volatilities that decline exponentially with their time horizons:

$$\sigma(t,T) = \sigma_r e^{-\alpha(T-t)} \tag{17.48}$$

where σ_r and α are positive constants. From (17.38), this implies that the rate of return volatility of a zero-coupon bond equals

$$\sigma_I(t,T) \equiv \int_t^T \sigma(t,s)\,ds = \int_t^T \sigma_r e^{-\alpha(s-t)} ds = \frac{\sigma_r}{\alpha}\left(1 - e^{-\alpha(T-t)}\right) \tag{17.49}$$

Note that this volatility function is the same as the Vasicek model of the term structure given in (9.44). Hence, the bond price's risk-neutral process is

claims cannot be valued by solving an equilibrium partial differential equation, such as was done in Chapter 9 in equation (9.40).

$dP(t,T)/P(t,T) = r(t)\,dt - \frac{\sigma_r}{\alpha}\left(1 - e^{-\alpha(T-t)}\right)d\widehat{z}$. To value contingent claims for this case, it remains to derive the instantaneous-maturity interest rate and its dynamics. From (17.45) and (17.46), we have

$$r(t) = f(0,t) + \int_0^t \frac{\sigma_r^2}{\alpha}\left(e^{-\alpha(t-u)} - e^{-2\alpha(t-u)}\right)du + \int_0^t \sigma_r e^{-\alpha(t-u)} d\widehat{z}(u) \quad (17.50)$$

$$\begin{aligned} dr &= \frac{\partial f(0,t)}{\partial t}dt + \int_0^t \left[\sigma_r^2 e^{-2\alpha(t-u)} - \sigma_r^2\left(e^{-\alpha(t-u)} - e^{-2\alpha(t-u)}\right)\right]dudt \\ &\quad - \int_0^t \alpha\sigma_r e^{-\alpha(t-u)} d\widehat{z}(u)\,dt + \sigma_r d\widehat{z} \end{aligned} \quad (17.51)$$

Substituting (17.50) into (17.51) and simplifying leads to

$$\begin{aligned} dr &= \frac{\partial f(0,t)}{\partial t}dt + \int_0^t \sigma_r^2 e^{-2\alpha(t-u)}dudt + \alpha\left[f(0,t) - r(t)\right]dt + \sigma_r d\widehat{z} \\ &= \alpha\left[\frac{1}{\alpha}\frac{\partial f(0,t)}{\partial t} + f(0,t) + \frac{\sigma_r^2}{2\alpha^2}\left(1 - e^{-2\alpha t}\right) - r(t)\right]dt + \sigma_r d\widehat{z} \\ &= \alpha\left[\overline{r}(t) - r(t)\right]dt + \sigma_r d\widehat{z} \end{aligned} \quad (17.52)$$

where $\overline{r}(t) \equiv \frac{1}{\alpha}\partial f(0,t)/\partial t + f(0,t) + \sigma_r^2\left(1 - e^{-2\alpha t}\right)/\left(2\alpha^2\right)$ is the risk-neutral central tendency of the short-rate process that is a deterministic function of time. The process in (17.52) is Markov in that the only stochastic variable affecting its future distribution is the current level of $r(t)$. However, it differs from the standard Vasicek model, which assumes that the risk-neutral process for $r(t)$ has a long-run mean that is constant.[19] By making the central tendency, $\overline{r}(t)$, a particular deterministic function of the currently observed forward rate curve, $f(0,t)\ \forall t \geq 0$, the model's implied date 0 price of a zero-coupon bond, $P(0,T)$, coincides exactly with observed prices.[20] This model was proposed by John

[19] Recall from equation (10.66) that the unconditional mean of the risk-neutral interest rate is $\overline{r} + q\sigma_r/\alpha$, where $\overline{r}$ is the mean of the physical process and q is the market price of interest rate risk.

[20] It is left as an exercise to verify that when $\overline{r}(t) \equiv \frac{1}{\alpha}\partial f(0,t)/\partial t + f(0,t) +$

Hull and Alan White ((Hull and White 1990); (Hull and White 1993)) and HJM (Heath, Jarrow, and Morton 1992) and is referred to as the "extended Vasicek" model.[21]

Let us illustrate this Extended Vasicek model by valuing a European option maturing at date T, where the underlying asset is a zero-coupon bond maturing at date $T+\tau$. Since, as with the standard Vasicek model, the extended Vasicek model has bond return volatilities as a deterministic function of time, the expectation in (17.47) for the case of a European option has an analytic solution. Alternatively, the results of Merton (Merton 1973b) given in equations (9.58) to (9.60) on the pricing of options when interest rates are random can be applied to derive the solution. However, instead of Chapter 9's assumption of the underlying asset being an equity that follows geometric Brownian motion, the underlying asset is a bond that matures at date $T+\tau$. For a call option with exercise price X, the boundary condition is $c(T) = \max[P(T, T+\tau) - X, 0]$. This leads to the solution

$$\begin{aligned} c(t) &= P(t, T+\tau)\ N(d_1) - P(t, T)\, X N(d_2) \\ &= e^{-\int_t^{T+\tau} f(t,s)ds} N(d_1) - e^{-\int_t^{T} f(t,s)ds} X N(d_2) \end{aligned} \tag{17.53}$$

where $d_1 = \left[\ln[P(t, T+\tau) / (P(t, T)\, X)] + \frac{1}{2} v(t, T)^2\right] / v(t, T)$, $d_2 = d_1 -$

$\sigma_r^2 \left(1 - e^{-2\alpha t}\right) / \left(2\alpha^2\right)$, then $P(0, T) = \widehat{E}\left[\exp\left(-\int_0^T r(s)\, ds\right)\right] = \exp\left(-\int_0^T f(0, s)\, ds\right)$.

[21] Hull and White show that, besides $\overline{r}(t)$, the parameters $\alpha(t)$ and $\sigma_r(t)$ also can be extended to be deterministic functions of time. With these extensions, $r(t)$ remains normally distributed and analytic solutions to options on discount bonds can be obtained. Making $\alpha(t)$ and $\sigma_r(t)$ time varying allows one to fit other aspects of the term structure, such as observed volatilities of forward rates.

$v(t,T)$, and where[22]

$$\begin{aligned} v(t,T)^2 &= \int_t^T \left[\sigma_I^2(t,u+\tau) + \sigma_I^2(t,u) - 2\rho\sigma_I(t,u+\tau)\sigma_I(t,u)\right] du \\ &= \frac{\sigma_r^2}{2\alpha^3}\left(1 - e^{-2\alpha(T-t)}\right)\left(1 - e^{-\sigma\tau}\right)^2 \end{aligned} \tag{17.54}$$

This solution illustrates a general principle of the HJM approach, namely, that formulas can be derived whose inputs match the initial term structure of bond prices ($P(t,T)$ and $P(t,T+\tau)$) or, equivalently, the initial forward rate curve ($f(t,s)\,\forall s \geq t$).

Our second example of a Markov HJM model is due to Peter Ritchken and L. Sankarasubramanian (Ritchken and Sankarasubramanian 1995), hereafter referred to as RS. They give general conditions on forward rate volatilities that result in term structure dynamics being Markov in two state variables. A particular example that satisfies these conditions is their example where forward rate volatilities take the form

$$\sigma(t,T) = \sigma_r r(t)^\gamma e^{-\alpha(T-t)} \tag{17.55}$$

where σ_r and α are positive constants. Thus, (17.55) specifies that the volatility of the short rate (when $T = t$) equals $\sigma_r r(t)^\gamma$. When $\gamma = 0$, we have our first example's extended Vasicek case of deterministic forward rates. However, empirical evidence indicates that interest rate volatility increases with the level of the short rate, so that it is desirable to obtain a Markov model with $\gamma > 0$. Similar to the derivation for $r(t)$ given for the extended Vasicek model, RS show that in this case the risk-neutral process for the instantaneous-maturity interest

[22] Note that when applying Merton's derivation to the case of the underlying asset being a bond, then ρ, the return correlation between bonds maturing at dates T and $T+\tau$, equals 1. This is because there is a single Brownian motion determining the stochastic component of returns.

rate satisfies

$$dr(t) = \left(\alpha\left[f(0,t) - r(t)\right] + \phi(t) + \frac{\partial f(0,t)}{\partial t}\right) dt + \sigma_r r(t)^{\gamma} d\widehat{z} \tag{17.56}$$

where

$$\begin{aligned}\phi(t) &= \int_0^t \sigma^2(s,t)\, ds \\ &= \sigma_r^2 \int_0^t r(s)^{2\gamma} e^{-2\alpha(t-s)} ds \end{aligned} \tag{17.57}$$

Differentiating (17.57) with respect to t, one obtains the dynamics of $\phi(t)$ to be

$$d\phi(t) = \left(\sigma_r^2 r(t)^{2\gamma} - 2\alpha\phi(t)\right) dt \tag{17.58}$$

The variable $\phi(t)$ is an "integrated variance" factor that evolves stochastically when $\gamma \neq 0$.[23] It, along with the short rate, $r(t)$, are two state variables that determine the evolution of $r(t)$. In turn, this determines the bonds' risk-neutral processes. Recall that since a bond's rate of return volatility equals $\sigma_I(t,T) \equiv \int_t^T \sigma(t,s)\, ds = \int_t^T \sigma_r r(t)^{\gamma} e^{-\alpha(s-t)} ds = \frac{\sigma_r r(t)^{\gamma}}{\alpha}\left(1 - e^{-\alpha(T-t)}\right)$, its risk-neutral price process equals

$$dP(t,T)/P(t,T) = r(t)\, dt - \frac{\sigma_r r(t)^{\gamma}}{\alpha}\left(1 - e^{-\alpha(T-t)}\right) d\widehat{z} \tag{17.59}$$

It is noteworthy that even when $\gamma = \frac{1}{2}$, the model differs from the CIR equilibrium model. Even though in both models the short rate's volatility, $\sigma_r\sqrt{r(t)}$, is the same, the RS model's requirement that it fit the observed term

[23] Note that when $\gamma = 0$, one obtains $\phi(t) = \frac{\sigma_r^2}{2\alpha}\left(1 - e^{-2\alpha t}\right)$, so that $\phi(t)$ is deterministic and the short rate process in (17.56) equals that of the extended Vasicek model in (17.52).

structure introduces a second stochastic state variable, $\phi(t)$, into the drift of the short rate process in (17.56).

In general, valuing fixed-income derivatives using the RS model does not lead to closed-form solutions. However, RS (Ritchken and Sankarasubramanian 1995) show that the risk-neutral processes for $r(t)$ and $\phi(t)$ can be discretized and Monte Carlo simulations performed to value contingent claims based on (17.47).

There are a number of other discrete-time models that can numerically value fixed-income derivatives based on calculations using binomial trees or lattices. These models can be viewed as discrete-time implementations of the continuous-time HJM approach in that they are designed to fit the initial term structure of bond prices and, possibly, bond volatilities. Thomas Ho and Sang Bin Lee (Ho and Lee 1986) first introduced the concept of pricing fixed-income derivatives by taking the initial term structure of bond prices as given and then making assumptions regarding the risk-neutral distribution of future interest rates. Their model is the discrete-time counterpart of the extended Vasicek model but with the mean reversion parameter, α, set to zero.[24] This binomial approach was modified for different risk-neutral interest rate dynamics by Fischer Black, Emanuel Derman, and William Toy (Black, Derman, and Toy 1990) and Fischer Black and Piotr Karasinski (Black and Karasinski 1991). These discrete-time "no-arbitrage" models are fixed-income counterparts to the binomial model of Chapter 7 that was used to price equity derivatives.

17.2.2 Market Models

[24]Thus, with zero mean reversion, an unattractive feature of this model is that the short rate is expected to explode over time. The Ho-Lee model is a mechanical way of calibrating a lattice that is consistent with an initial term structure of bond prices. The HJM approach can be viewed as a shortcut to accomplishing this because the extended Vasicek model provides an analytic solution that embeds the Ho-Lee assumptions.

As shown in the previous section, HJM models begin with a particular specification for instantaneous-maturity, continuously compounded forward rates, and then derivative values are calculated based on these initial forward rates. However, instantaneous-maturity forward rates are not directly observable, and in many applications they must be approximated from data on bond yields or discrete-maturity forward or futures rates that are unavailable at every maturity. A class of models that is a variation on the HJM approach can sometimes avoid this approximation error and may lead to more simple, analytic solutions for particular types of derivatives. These models are known as "market models" and are designed to price derivatives whose payoffs are a function of a discrete maturity, rather than instantaneous-maturity, forward interest rate. Examples of such derivatives include interest rate caps and floors and swaptions. Let us illustrate the market model approach by way of these examples.

Example: An Interest Rate Cap

Consider valuing a European option written on a discrete forward rate, such as one based on the London Interbank Offer Rate (LIBOR). Define $L(t,T,\tau)$ as the date t annualized, τ-period compounded, forward interest rate for borrowing or lending over the period from future date T to $T+\tau$.[25] In terms of current date t discount bond prices $(P(t,t+\delta))$, forward price $(F(t,T,\tau))$, and continuously compounded forward rate $(f(t,T,\tau))$, this discrete forward rate is defined by the relation

$$\frac{P(t,T+\tau)}{P(t,T)} = F(t,T,\tau) = e^{-f(t,T,\tau)\tau} = \frac{1}{1+\tau L(t,T,\tau)} \tag{17.60}$$

[25] The convention for LIBOR is to set the compounding interval equal to the underlying instrument's maturity. For example, if $\tau = \frac{1}{4}$ years, then three-month LIBOR is compounded quarterly If $\tau = \frac{1}{2}$ years, then six-month LIBOR is compounded semiannually.

Note that when $T = t$, $P(t, t+\tau) = 1/[1 + \tau L(t, t, \tau)]$ defines $L(t, t, \tau)$ as the current "spot" τ-period LIBOR.[26] An example of an option written on LIBOR is a caplet that matures at date $T + \tau$ and is based on the realized spot rate $L(T, T, \tau)$. Assuming this caplet has an exercise cap rate X, its date $T + \tau$ payoff is

$$c(T+\tau) = \tau \max[L(T, T, \tau) - X, 0] \tag{17.61}$$

that is, the option payoff at date $T + \tau$ depends on the τ-period spot LIBOR at date T.[27] Because uncertainty regarding the LIBOR rate is resolved at date T, which is τ period's prior to the caplet's settlement (payment) date, we can also write

$$\begin{aligned} c(T) &= P(T, T+\tau) \max[\tau L(T, T, \tau) - \tau X, 0] \\ &= P(T, T+\tau) \max\left[\frac{1}{P(T, T+\tau)} - 1 - \tau X, 0\right] \\ &= \max[1 - (1+\tau X) P(T, T+\tau), 0] \\ &= \max\left[1 - \frac{1+\tau X}{1+\tau L(T, T, \tau)}, 0\right] \end{aligned} \tag{17.62}$$

which illustrates that a caplet maturing at date $T + \tau$ is equivalent to a put option that matures at date T, has an exercise price of 1, and is written on a zero-coupon bond that has a payoff of $1 + \tau X$ at its maturity date of $T + \tau$. Similarly, a floorlet, whose date $T + \tau$ payoff equals $\tau \max[X - L(T, T, \tau), 0]$, can be shown to be equivalent to a call option on a zero-coupon bond.[28]

To value a caplet using a market model approach, let us first analyze the

---

[26] This modeling assumes that LIBOR is the yield on a default-free discount bond. However, LIBOR is not a fully default-free interest rate, such as a Treasury security rate. It represents the borrowing rate of a large, generally high-credit-quality, bank. Typically, the relatively small amount of default risk is ignored when applying market models to derivatives based on LIBOR.

[27] Caplets are based on a notional principal amount, which here is assumed to be \$1. The value of a caplet having a notional principal of \$$N$ is simply N times the value of a caplet with a notional principal of \$1, that is, its payoff is $\tau N \max[L(T, T, \tau) - X, 0]$.

[28] Therefore, the HJM-extended Vasicek solution in (17.53) to (17.54) is one method for valuing a floorlet. A straightforward modification of this formula could also value a caplet.

dynamics of $L(t,T,\tau)$. Rearranging (17.60) gives

$$\tau L(t,T,\tau) = \frac{P(t,T)}{P(t,T+\tau)} - 1 \tag{17.63}$$

We can derive the stochastic process followed by this forward rate in terms of the bond prices' risk-neutral processes. Note that from (17.38), along with $\mathbf{dz} = \mathbf{d\widehat{z}} - \mathbf{\Theta}(t)\,dt$, we have $dP(t,T)/P(t,T) = r(t)\,dt - \boldsymbol{\sigma}_I(t,T)'\,\mathbf{d\widehat{z}}$. Applying Itô's lemma to (17.63), we obtain

$$\begin{aligned}\frac{dL(t,T,\tau)}{L(t,T,\tau)} &= \left(\boldsymbol{\sigma}_I(t,T+\tau)'\left[\boldsymbol{\sigma}_I(t,T+\tau) - \boldsymbol{\sigma}_I(t,T)\right]\right)dt \\ &\quad + \left[\boldsymbol{\sigma}_I(t,T+\tau) - \boldsymbol{\sigma}_I(t,T)\right]'\mathbf{d\widehat{z}}\end{aligned} \tag{17.64}$$

In principle, now we could value a contingent claim written on $L(t,T,\tau)$ by calculating the claim's discounted expected terminal payoff assuming $L(t,T,\tau)$ follows the process in (17.64).[29] However, as will become clear, there is an alternative probability measure to the one generated by $\mathbf{d\widehat{z}}$ that can be used to calculate a contingent claim's expected payoff, and this alternative measure is analytically more convenient for this particular forward rate application.

To see this, consider the new transformation $\mathbf{d\widetilde{z}} = \mathbf{d\widehat{z}} + \boldsymbol{\sigma}_I(t,T+\tau)\,dt = \mathbf{dz} + \left[\mathbf{\Theta}(t) + \boldsymbol{\sigma}_I(t,T+\tau)\right]dt$. Substituting into (17.64) results in

$$\frac{dL(t,T,\tau)}{L(t,T,\tau)} = \left[\boldsymbol{\sigma}_I(t,T+\tau) - \boldsymbol{\sigma}_I(t,T)\right]'\mathbf{d\widetilde{z}} \tag{17.65}$$

so that under the probability measure generated by $\mathbf{d\widetilde{z}}$, the process followed by $L(t,T,\tau)$ is a martingale. This probability measure is referred to as the *forward rate measure* at date $T+\tau$. Note that since $L(t,T,\tau)$ is linear in the

[29] Specifically, if $c(t,L(t,T,\tau))$ is the contingent claim's value, it could be calculated as $\widehat{E}_t\left[e^{-\int_t^T r(s)ds}c(T,L(T,T,\tau))\right]$ where $r(t)$ and $L(t,T,\tau)$ are assumed to follow risk-neutral processes.

bond price $P(t,T)$ deflated by $P(t,T+\tau)$, the forward rate measure at date $T+\tau$ works by deflating all security prices by the price of the discount bond that matures at date $T+\tau$. This contrasts with the risk-neutral measure where security prices are deflated by the value of the money market account, which follows the process $dB(t) = r(t)B(t)\,dt$.

Not only does $L(t,T,\tau)$ follow a martingale under the forward measure, but so does the value of all other securities. To see this, let the date t price of a contingent claim be given by $c(t)$. In the absence of arbitrage, its price process is of the form

$$\frac{dc}{c} = [r(t) + \boldsymbol{\Theta}(t)'\boldsymbol{\sigma}_c(t)]\,dt + \boldsymbol{\sigma}_c(t)'\,\mathbf{dz} \tag{17.66}$$

Now define the deflated contingent claim's price as $C(t) = c(t)/P(t,T+\tau)$. Applying Itô's lemma gives

$$\begin{aligned}\frac{dC}{C} &= [\boldsymbol{\Theta}(t) + \boldsymbol{\sigma}_I(t,T+\tau)]'[\boldsymbol{\sigma}_c(t) + \boldsymbol{\sigma}_I(t,T+\tau)]\,dt \\ &\quad + [\boldsymbol{\sigma}_c(t) + \boldsymbol{\sigma}_I(t,T+\tau)]'\mathbf{dz}\end{aligned} \tag{17.67}$$

and making the forward measure transformation $\mathbf{d\widetilde{z}} = \mathbf{dz} + [\boldsymbol{\Theta}(t) + \boldsymbol{\sigma}_I(t,T+\tau)]\,dt$, (17.67) becomes the martingale process

$$\frac{dC}{C} = [\boldsymbol{\sigma}_c(t) + \boldsymbol{\sigma}_I(t,T+\tau)]'\mathbf{d\widetilde{z}} \tag{17.68}$$

so that $C(t) = \widetilde{E}_t[C(t+\delta)]\ \forall\delta \geq 0$, where $\widetilde{E}_t[\cdot]$ is the date t expectation under the forward measure. Now, to show why this transformation can be convenient, suppose that this contingent claim is the caplet described earlier. This deflated caplet's value is given by

$$\begin{aligned} C(t) &= \widetilde{E}_t\left[C(T+\tau)\right] \\ &= \widetilde{E}_t\left[\frac{\tau \max\left[L(T,T,\tau)-X,0\right]}{P(T+\tau,T+\tau)}\right] \end{aligned} \tag{17.69}$$

Noting that $C(t) = c(t)/P(t,T+\tau)$ and realizing that $P(T+\tau,T+\tau) = 1$, we can rewrite this as

$$c(t) = P(t,T+\tau)\,\widetilde{E}_t\left[\tau \max\left[L(T,T,\tau)-X,0\right]\right] \tag{17.70}$$

A common practice is to assume that $L(T,T,\tau)$ is lognormally distributed under the date $T+\tau$ forward measure.[30] This means that $[\boldsymbol{\sigma}_I(t,T+\tau) - \boldsymbol{\sigma}_I(t,T)]$ in (17.65) must be a vector of nonstochastic functions of time that can be calibrated to match observed bond or forward rate volatilities.[31] Noting that $L(t,T,\tau)$ also has a zero drift leads to a similar formula first proposed by Fischer Black (Black 1976) for valuing options on commodity futures:

$$c(t) = \tau P(t,T+\tau)\left[L(t,T,\tau)N(d_1) - XN(d_2)\right] \tag{17.71}$$

where $d_1 = \left[\ln\left(L(t,T,\tau)/X\right) + \frac{1}{2}v(t,T)^2\right]/v(t,T)$, $d_2 = d_1 - v(t,T)$, and

$$v(t,T)^2 = \int_t^T \left|\boldsymbol{\sigma}_I(s,T+\tau) - \boldsymbol{\sigma}_I(s,T)\right|^2 ds \tag{17.72}$$

Equation (17.71) is similar to equation (10.60) derived in Chapter 10 for the case

[30] Assuming a lognormal distribution for $L(t,T,\tau)$ is attractive because it prevents this discrete forward rate from becoming negative, thereby also restricting yields on discount bonds to be nonnegative. Note that if instantaneous-maturity forward rates are assumed to be lognormally distributed, HJM show that they will be expected to become infinite in finite time. This is inconsistent with arbitrage-free bond prices. Fortunately, such an explosion of rates does not occur when forward rates are discrete (Brace, Gatarek, and Musiela 1997).

[31] Note that since $\boldsymbol{\sigma}_I(t,t+\delta)$ is an integral of instantaneous forward rate volatilities, the lognormality of $\boldsymbol{\sigma}_I(t,t+\tau) - \boldsymbol{\sigma}_I(t,T)$ puts restrictions on instantaneous forward rates under an HJM modeling approach. However, we need not focus on this issue for pricing applications involving a discrete forward rate.

of a call option on a forward or futures price where the underlying is lognormally distributed and interest rates are nonstochastic.

An interest rate cap is a portfolio of caplets written on the same τ-period LIBOR but maturing at different dates $T = T_1, T_2, ..., T_n$, where typically $T_{j+1} = T_j + \tau$. Standard practice is to value each individual caplet in the portfolio in the manner we have described, where the caplet maturing at date T_j is priced using the date $T_j + \tau$ forward measure. Often, caps are purchased by issuers of floating-rate bonds whose bond payments coincide with the caplet maturity dates. Doing so insures the bond issuer against having to make a floating coupon rate greater than X (plus a credit spread). Since a floating-rate bond's coupon rate payable at date $T + \tau$ is most commonly tied to the τ-period LIBOR at date T, caplet payoffs follow this same structure. Analogous to a cap, an interest rate floor is a portfolio of floorlets and can be valued using the same technique described in this section.

Example: A Swaption

Frequently, a market model approach is applied to value another common interest rate derivative, a swaption. A *swaption* is an option to become a party in an interest rate swap at a given future maturity date and at a prespecified swap rate. Let us, then, define the interest rate swap underlying this swaption. A standard "plain vanilla" swap is an agreement between two parties to exchange fixed interest rate coupon payments for floating interest rate coupon payments at dates $T_1, T_2, ..., T_{n+1}$, where $T_{j+1} = T_j + \tau$ and τ is the maturity of the LIBOR of the floating-rate coupon payments. Thus, if K is the swap's fixed annualized coupon rate, then at date T_{j+1} the fixed-rate payer's net payment is $\tau [K - L(T_j, T_j, \tau)]$, whereas that of the floating-rate payer is exactly the opposite.[32]

Note that the swap's series of floating-rate payments plus an additional \$1 at date T_{n+1} can be replicated by starting with \$1 at time $T_0 = T_1 - \tau$ and repeatedly investing this \$1 in τ-maturity LIBOR deposits.[33] These are the same cashflows that one would obtain by investing \$1 in a floating-rate bond at date T_0. Similarly, the swap's series of fixed-rate payments plus an additional \$1 at date T_{n+1} can be replicated by buying a fixed-coupon bond that pays coupons of τK at each swap date and pays a principal of \$1 at its maturity date of T_{n+1}. Based on this insight, one can see that the value of a swap to the floating-rate payer is the difference between a fixed-coupon bond having coupon rate K, and a floating-coupon bond having coupons tied to τ-period LIBOR. Thus, if $t \leq T_0 = T_1 - \tau$, then the date t value of the swap to the floating-rate payer is[34]

$$\tau K \sum_{j=1}^{n+1} P(t, T_j) + P(t, T_{n+1}) - P(t, T_0) \tag{17.73}$$

When a standard swap agreement is initiated at time T_0, the fixed rate K is set such that the value of the swap in (17.73) is zero. This concept of setting K to make the agreement fair (similar to forward contracts) can be extended to dates prior to T_0. One can define $s_{0,n}(t)$ as the forward swap rate that makes the date t value of the swap (starting at date T_0 and making n subsequent exchanges) equal to zero. Setting $K = s_{0,n}(t)$ and equating (17.73) to zero,

[32] Recall that $L(T_j, T_j, \tau)$ is the spot τ-period LIBOR at date T_j. Also, as discussed in the preceding footnote, this exchange is based on a notional principal of \$1. For a notional principal of \$$N$, all payments are multiplied by N.

[33] Thus, \$1 invested at time T_0 produces a return of $1 + \tau L(T_0, T_0, \tau)$ at T_1. Keeping the cashflow of $\tau L(T_0, T_0, \tau)$ and reinvesting the \$1 will then produce a return of $1+\tau L(T_1, T_1, \tau)$ at T_2. Keeping the cashflow of $\tau L(T_1, T_1, \tau)$ and reinvesting the \$1 will then produce a return of $1 + \tau L(T_2, T_2, \tau)$ at T_3. This process is repeated until at time T_{n+1} a final return of $1 + \tau L(T_n, T_n, \tau)$ is obtained.

[34] Notice that $P(t, T_0)$ is the date t value of the floating-rate bond while the remaining terms are the value of the fixed-rate bond.

one obtains

$$s_{0,n}(t) = \frac{P(t,T_0) - P(t,T_{n+1})}{\tau \sum_{j=1}^{n+1} P(t,T_j)} = \frac{P(t,T_0) - P(t,T_{n+1})}{B_{1,n}(t)} \tag{17.74}$$

where $B_{1,n}(t) \equiv \tau \sum_{j=1}^{n+1} P(t,T_j)$ is a portfolio of zero-coupon bonds that each pay τ at the times of the swap's exchanges.

Now a standard swaption is an option to become either a fixed-rate payer or floating-rate payer at a fixed swap rate X at a specified future date. Thus, if the maturity of the swaption is date T_0, at which time the holder of the swaption has the right but not the obligation, to become a fixed-rate payer (floating-rate receiver), this option's payoff equals[35]

$$c(T_0) = \max\left[B_{1,n}(T_0)\left[s_{0,n}(T_0) - X\right], 0\right] = \max\left[1 - P(T_0,T_{n+1}) - B_{1,n}(T_0)X, 0\right] \tag{17.75}$$

Note from the first line of (17.75) that when the option is in the money, then $B_{1,n}(T_0)\left[s_{0,n}(T_0) - X\right]$ is the date T_0 value of the fixed-rate payer's savings from having the swaption relative to entering into a swap at the fair spot rate $s_{0,n}(T_0)$. In the second line of (17.75), we have substituted from (17.74) $s_{0,n}(T_0)B_{1,n}(T_0) = P(T_0,T_0) - P(T_0,T_{n+1}) = 1 - P(T_0,T_{n+1})$. This illustrates that a swaption is equivalent to an option on a coupon bond with coupon rate X and an exercise price of 1.

To value this swaption at date $t \leq T_0$, a convenient approach is to recognize from (17.75) that the swaption's payoff is proportional to $B_{1,n}(T_0) \equiv \tau \sum_{j=1}^{n+1} P(T_0,T_j)$. This suggests that $B_{1,n}(t)$ is a convenient deflator for valu-

[35] The payoff of an option to be a floating-rate payer (fixed-rate receiver) is $\max\left[B_{1,n}(T_0)\left[X - s_{0,n}(T_0)\right], 0\right]$.

ing the swap. By normalizing all security prices by $B_{1,n}(t)$, we will value the swaption using the so-called "forward swap measure."

Similar to valuation under the risk-neutral or forward measure of the previous section, let us define $C(t) = c(t)/B_{1,n}(t)$. Also define $\mathbf{d\overline{z}} = \mathbf{dz} + \left[\mathbf{\Theta}(t) + \boldsymbol{\sigma}_{B_{1,n}}(t)\right] dt$ where $\boldsymbol{\sigma}_{B_{1,n}}(t)$ is the date t vector of instantaneous volatilities of the zero-coupon bond portfolio's value, $B_{1,n}(t)$. Similar to the derivation in equations (17.66) to (17.68), we have

$$\frac{dC}{C} = \left[\boldsymbol{\sigma}_c(t) + \boldsymbol{\sigma}_{B_{1,n}}(t, T+\tau)\right]' \mathbf{d\overline{z}} \tag{17.76}$$

so that all deflated asset prices under the forward swap measure follow martingale processes. Thus,

$$\begin{aligned} C(t) &= \overline{E}_t\left[C(T_0)\right] \qquad (17.77) \\ &= \overline{E}_t\left[\frac{\max\left[B_{1,n}(T_0)\left[s_{0,n}(T_0) - X\right], 0\right]}{B_{1,n}(T_0)}\right] \\ &= \overline{E}_t\left[\max\left[s_{0,n}(T_0) - X, 0\right]\right] \end{aligned}$$

Rewritten in terms of the undeflated swaption's current value, $c(t) = C(t) B_{1,n}(t)$, (17.77) becomes

$$c(t) = B_{1,n}(t)\, \overline{E}_t\left[\max\left[s_{0,n}(T_0) - X, 0\right]\right] \tag{17.78}$$

so that the expected payoff under the forward swap measure is discounted by the current value of a portfolio of zero-coupon bonds that mature at the times of the swap's exchanges.

Importantly, note that $s_{0,n}(t) = \left[P(t, T_0) - P(t, T_{n+1})\right]/B_{1,n}(t)$ is the ratio of the difference between two security prices deflated by $B_{1,n}(t)$. In the absence of arbitrage, it must also follow a martingale process under the forward swap measure. A convenient and commonly made assumption is that this forward swap rate is lognormally distributed under the forward swap measure:

$$\frac{ds_{0,n}(t)}{s_{0,n}(t)} = \boldsymbol{\sigma}_{s_{0,n}}(t)' \mathbf{d\bar{z}} \tag{17.79}$$

so that $\boldsymbol{\sigma}_{S_{0,n}}(t)$ is a vector of deterministic functions of time that can be calibrated to match observed forward swap volatilities or zero-coupon bond volatilities.[36] This assumption results in (17.78) taking a Black-Scholes-type form:

$$c(t) = B_{1,n}(t)\left[s_{0,n}(t)N(d_1) - XN(d_2)\right] \tag{17.80}$$

where $d_1 = \left[\ln(s_{0,n}(t)/X) + \frac{1}{2}v(t,T_0)^2\right]/v(t,T_0)$, $d_2 = d_1 - v(t,T_0)$, and

$$v^2(t,T_0) = \int_t^{T_0} \boldsymbol{\sigma}_{s_{0,n}}(u)' \boldsymbol{\sigma}_{s_{0,n}}(u)\, du \tag{17.81}$$

17.2.3 Random Field Models

The term structure models that we have studied thus far have specified a finite number of Brownian motion processes as the source of uncertainty determining the evolution of bond prices or forward rates. For example, the bond price processes in equilibrium models (see equation (17.2)) and HJM models (see equation (17.38)) were driven by an $n \times 1$ vector of Brownian motions, $\mathbf{dz}$. One implication of this is that a Black-Scholes hedge portfolio of n different maturity bonds can be used to perfectly replicate the risk of any other maturity bond. As shown in Chapter 10, in the absence of arbitrage, the fact that any bond's risk can be hedged with other bonds places restrictions on bonds' expected excess rates of return and results in a unique vector of market prices of risk, $\boldsymbol{\Theta}(t)$,

[36] Applying Itô's lemma to (17.74) allows one to derive the volatility of $s_{0,n}(t)$ in terms of zero-coupon bond volatilities.

associated with $\mathbf{dz}$. This implies a bond price process of the form

$$dP(t,T)/P(t,T) = \left[r(t) + \mathbf{\Theta}(t)' \boldsymbol{\sigma}_p(t,T)\right] dt + \boldsymbol{\sigma}_p(t,T)' \mathbf{dz} \qquad (17.82)$$

Moreover, the Black-Scholes hedge, by making the market dynamically complete and by identifying a unique $\mathbf{\Theta}(t)$ associated with $\mathbf{dz}$, allows us to perform risk-neutral valuation by the transformation $\mathbf{d\widehat{z}} = \mathbf{dz} + \mathbf{\Theta}(t)\, dt$ or valuation using the pricing kernel $dM/M = -r(t)\, dt - \mathbf{\Theta}(t)' \mathbf{dz}$.

However, the elegance of these models comes with an empirical downside. The fact that all bond prices depend on the same $n \times 1$ vector $\mathbf{dz}$ places restrictions on the covariance of bonds' rates of return. For example, when $n = 1$, the rates of return on all bonds are instantaneously perfectly correlated. While in these models the correlation can be made less perfect by increasing n, doing so introduces more parameters that require estimation.

A related empirical implication of (17.82) or (17.35) is that it restricts the possible future term structures of bond prices or forward rates. In other words, starting from the current date t set of bond prices $P(t,T)\ \forall T > t$, an arbitrary future term structure, $P(t+dt,T)\ \forall T > t+dt$, cannot always be achieved by any realization of $\mathbf{dz}$. This is because a given future term structure has an infinite number of bond prices (each of a different maturity), but the finiteness of $\mathbf{dz}$ allows matching this future term structure at only a finite number of maturity horizons.[37] Hence, models based on a finite $\mathbf{dz}$ are almost certainly inconsistent with future observed bond prices and forward rates. Because of this, empiricists must assume that data on bond prices (or yields) are observed with "noise" or that, in the case of HJM-type models, parameters (that the model assumes to be constant) must be recalibrated at each observation date

[37] For example, consider $n = 1$. In this case, all bond prices must either rise or fall with a given realization of dz. This model would not permit a situation where short-maturity bond prices fell but long-maturity bond prices rose. The model could produce a realization of dz that matched long-maturity bond prices or short-maturity bond prices, but not both.

to match the new term structure of forward rates.

Random field models are an attempt to avoid these empirical deficiencies. Research in this area includes that of David Kennedy (Kennedy 1994); (Kennedy 1997), Robert Goldstein (Goldstein 2000), Pedro Santa-Clara and Didier Sornette (Santa-Clara and Sornette 2001), and Robert Kimmel (Kimmel 2004). These models specify that each zero-coupon bond price, $P(t,T)$, or each instantaneous forward rate, $f(t,T)$, is driven by a Brownian motion process that is unique to the bond's or rate's maturity, T. For example, a model of this type might assume that a bond's risk-neutral process satisfies

$$dP(t,T)/P(t,T) = r(t)\,dt + \sigma_p(t,T)\,d\widehat{z}_T \qquad \forall T > t \tag{17.83}$$

where $d\widehat{z}_T(t)$ is a single Brownian motion process (under the risk-neutral measure) that is unique to the bond that matures at date T.[38] The set of Brownian motions for all zero-coupon bonds $\{\ \widehat{z}_T(t)\ \}_{T>t}$ comprises a Brownian "field," or "sheet." This continuum of Brownian motions has two dimensions: calendar time, t, and time to maturity, T. The elements affecting different bonds are linked by an assumed correlation structure:

$$d\widehat{z}_{T_1}(t)\,d\widehat{z}_{T_2}(t) = \rho(t,T_1,T_2)\,dt \tag{17.84}$$

where $\rho(t,T_1,T_2) > 0$ is specified to be a particular continuous, differentiable function with $\rho(t,T,T) = 1$ and $\frac{\partial\rho(t,T_1,T_2)}{\partial T_1}|_{T_1=T_2} = 0$. For example, one simple specification involving only a single parameter is $\rho(t,T_1,T_2) = e^{-\varrho|T_1-T_2|}$, where ϱ is a positive constant.

One can also model the physical process for bond prices corresponding to

[38] An alternative way of specifying a random field model is to assume that the risk-neutral processes for instantaneous forward rates are of the form $df(t,T) = \left[\sigma(t,T)\int_t^T \sigma(t,s)\,c(t,T,s)\,ds\right]dt + \sigma(t,T)\,dz_T$, where $dz_{T_1}dz_{T_2} = c(t,T_1,T_2)\,dt$. This specification extends the HJM equation (17.43) to a random field driving forward rates.

(17.83). If $\theta_T(t)$ is the market price of risk associated with $d\widehat{z}_T(t)$, then making the transformation $dz_T = d\widehat{z}_T + \theta_T(t)\,dt$, one obtains

$$dP(t,T)/P(t,T) = [r(t) + \theta_T(t)\,\sigma_p(t,T)]\,dt + \sigma_p(t,T)\,dz_T \qquad \forall T > t \quad (17.85)$$

with $dz_T(t)$, $T > t$ satisfying the same correlation function as in (17.84). Analogous to the finite-factor pricing kernel process in (17.8), a pricing kernel for this random field model would be

$$dM/M = -r(t)\,dt - \int_t^{\infty} [\theta_T(t)\,dz_T(t)]\,dT \qquad (17.86)$$

so that an integral of the products of market prices of risk and Brownian motions replaces the usual sum of these products that occur for the finite factor case.[39]

The benefit of a model like (17.83) and (17.84) is that a realization of the Brownian field can generate any future term structure of bond prices or forward rates and, hence, be consistent with empirical observation and not require model recalibration. Moreover, with only a few additional parameters, random field models can provide a flexible covariance structure among different maturity bonds. Specifically, unlike finite-dimensional equilibrium models or HJM models, the covariance matrix of different maturity bond returns or forward rates will always be nonsingular no matter how many bonds are included. This could be important when valuing particular fixed-income derivatives where the underlying is a portfolio of zero-coupon bonds, and the correlation between these bonds affects the overall portfolio volatility.

[39] Note, however, that a random field model is not the same as a standard finite factor model extended to an infinite number of factors. As shown in (17.85), a random field model has a single Brownian motion driving each bond price or forward rate. A factor model, such as (17.2) or (17.38), extended to infinite factors would have the same infinite set of Brownian motions driving each bond price.

However, this rich covariance structure requires stronger theoretical assumptions for valuing derivatives compared to finite-dimensional diffusion models. A given bond's return can no longer be perfectly replicated by a portfolio of other bonds, and thus a Black-Scholes hedging argument cannot be used to identify a unique market price of risk associated with each $dz_T(t)$.[40] The market for fixed-income securities is no longer dynamically complete. Hence, one must assume, perhaps due to an underlying preference-based general equilibrium model, that there exists particular $\theta_T(t)$ associated with each $dz_T(t)$ or, equivalently, that a risk-neutral pricing exists.

Random field models can be parameterized by assuming particular functions for bond price or forward rate volatilities. For example, Pierre Collin-Dufresne and Robert Goldstein (Collin-Dufresne and Goldstein 2003) propose a stochastic volatility model where, in equation (17.83), $\sigma_p(t,T) = \sigma(t,T)\sqrt{\Sigma(t)}$, where $\sigma(t,T)$ is a deterministic function and where $\Sigma(t)$ is a volatility factor, common to all bonds, that follows the square root process

$$d\Sigma(t) = \kappa\left(\overline{\Sigma} - \Sigma(t)\right)dt + \vartheta\sqrt{\Sigma(t)}d\widehat{z}_\Sigma \tag{17.87}$$

where $d\widehat{z}_\Sigma$ is a Brownian motion (under the risk-neutral measure) that is assumed to be independent of the Brownian field $\{d\widehat{z}_T\}$ $\forall T > t$. Based on this parameterization, which is similar to a one-factor affine model, they derive solutions for various interest rate derivatives.[41]

[40] Robert Goldstein (Goldstein 2000) characterizes random field models of the term structure as being analogous to the APT model (Ross 1976). As discussed in Chapter 3, the APT assumes that a given asset's return depends on the risk from a finite number of factors along with the asset's own idiosyncratic risk. Thus, the asset is imperfectly correlated with any portfolio containing a finite number of other assets. Similarly, in a random field model, a given bond's return is imperfectly correlated with any portfolio containing a finite number of other bonds. Taking the analogy a step further, perhaps market prices of risk in a random field model can be characterized using the notion of asymptotic arbitrage, rather than exact arbitrage.

[41] Robert Kimmel (Kimmel 2004) also derives models with stochastic volatility driven by multiple factors.

If, similar to David Kennedy (Kennedy 1994), one makes the more simple assumption that $\sigma_p(t,T)$ in (17.83) and $\rho(t,T_1,T_2)$ in (17.84) are deterministic functions, then options on bonds, such as caplets and floorlets, have a Black-Scholes-type valuation formula. For example, suppose as in the HJM-extended Vasicek case of (17.53) to (17.54) that we value a European call option that matures at date T, is written on a zero-coupon bond that matures at date $T+\tau$, and has an exercise price of X. Similar to (17.70), we can value this option using the date T forward rate measure:

$$c(t) = P(t,T)\,\widetilde{E}_t\left[\max\left[p(T,T+\tau) - X, 0\right]\right] \tag{17.88}$$

where $p(t,T+\tau) \equiv P(t,T+\tau)/P(t,T)$ is the deflated price of the bond that matures at date $T+\tau$. Applying Itô's lemma to the risk-neutral process for bond prices in (17.83), we obtain

$$\begin{aligned}\frac{dp(t,T+\tau)}{p(t,T+\tau)} &= \sigma_p(t,T)\left[\sigma_p(t,T) - \rho(t,T,T+\tau)\,\sigma_p(t,T+\tau)\right]dt \\ &\quad +\sigma_p(t,T+\tau)\,d\widehat{z}_{T+\tau} - \sigma_p(t,T)\,d\widehat{z}_T \end{aligned} \tag{17.89}$$

We can rewrite $d\widehat{z}_{T+\tau} = \rho(t,T,T+\tau)\,d\widehat{z}_T + \sqrt{1-\rho(t,T,T+\tau)^2}\,d\widehat{z}_{U,T}$, where $d\widehat{z}_{U,T}$ is a Brownian motion uncorrelated with $d\widehat{z}_T$, so that the stochastic component in (17.89) can be written as $\sigma_p(t,T+\tau)\sqrt{1-\rho(t,T,T+\tau)^2}\,d\widehat{z}_{U,T}$ $+\left[\sigma_p(t,T+\tau)\,\rho(t,T,T+\tau) - \sigma_p(t,T)\right]d\widehat{z}_T$.[42] Then making the transformation to the date T forward measure, $d\widetilde{z}_T = d\widehat{z}_T + \sigma_p(t,T)$, the process for

[42] This rewriting puts the risk-neutral process for $p(t,T+\tau)$ in the form of our prior analysis in which the vector of Brownian motions, $\mathbf{d\widehat{z}}$, was assumed to have independent elements. This allows us to make the transformation to the forward measure in the same manner as was done earlier.

$p(t, T+\tau)$ becomes

$$\begin{aligned} \frac{dp(t, T+\tau)}{p(t, T+\tau)} &= \sigma_p(t, T+\tau)\sqrt{1-\rho(t, T, T+\tau)^2}\,d\widehat{z}_{U,T} \\ &\quad + \left[\sigma_p(t, T+\tau)\,\rho(t, T, T+\tau) - \sigma_p(t, T)\right] d\widetilde{z}_T \\ &= \sigma(t, T, \tau)\,d\widetilde{z} \end{aligned} \tag{17.90}$$

where

$$\sigma(t, T, \tau)^2 \equiv \sigma_p(t, T+\tau)^2 + \sigma_p(t, T)^2 - 2\rho(t, T, T+\tau)\,\sigma_p(t, T+\tau)\,\sigma_p(t, T) \tag{17.91}$$

Thus, $p(t, T+\tau)$ is lognormally distributed under the forward rate measure, so that (17.88) has the Black-Scholes-Merton-type solution

$$\begin{aligned} c(t) &= P(t, T)\left[p(t, T+\tau)\,N(d_1) - XN(d_2)\right] \qquad (17.92) \\ &= P(t, T+\tau)\,N(d_1) - P(t, T)\,N(d_2) \end{aligned}$$

where $d_1 = \left[\ln\left(p(t, T+\tau)/X\right) + \frac{1}{2}v(t, T)^2\right]/v(t, T)$, $d_2 = d_1 - v(t, T)$, and

$$v(t, T)^2 = \int_t^T \sigma(u, T, \tau)^2\,du \tag{17.93}$$

and $\sigma(u, T, \tau)$ is defined in (17.91). While this formula is similar to the Vasicek-based ones in (9.58) and (17.53), the volatility function in (17.91) may permit a relatively more flexible form for matching observed data.

17.3 Summary

This chapter has outlined some of the important theoretical developments in modeling bond yield curves and valuing fixed-income securities. The chapter's presentation has been in the context of continuous-time models and, to keep its length manageable, many similar models set in discrete time have been omitted.[43] Moreover, questions regarding numerical implementation and parameter estimation for specific models could not be addressed in the short presentations given here.

There is a continuing search for improved ways of describing the term structure of bond prices and of valuing fixed-income derivatives. Researchers in this field have different objectives, and the models that we presented reflect this diversity. Much academic research focuses on analyzing equilibrium models in hopes of better understanding the underlying macroeconomic factors that shape the term structure of bond yields. In contrast, practitioner research concentrates on models that can value and hedge fixed-income derivatives. Their ideal model would match the initial term structure, provide a parsimonious structure for forward rate volatilities, and avoid negative, exploding forward rates. Unfortunately, a model with all of these characteristics is hard to find.

While in recent years research on term structure models has expanded, studies in the related field of default-risky fixed-income securities have grown even more rapidly. The next chapter takes up this topic of valuing defaultable bonds and credit derivatives.

17.4 Exercises

1. Consider the following example of a two-factor term structure model (Jegadeesh and Pennacchi 1996); (Balduzzi, Das, and Foresi 1998). The instantaneous-

[43] Treatments of models set in discrete time include books by Robert Jarrow (Jarrow 2002), Bruce Tuckman (Tuckman 2002), and Thomas Ho and Sang Bin Lee (Ho and Lee 2004).

maturity interest rate is assumed to follow the physical process

$$dr(t) = \alpha \left[\gamma(t) - r(t)\right] dt + \sigma_r dz_r$$

and the physical process for the interest rate's stochastic "central tendency," $\gamma(t)$, satisfies

$$d\gamma(t) = \delta \left[\overline{\gamma} - \gamma(t)\right] dt + \sigma_\gamma dz_\gamma$$

where $dz_r dz_\gamma = \rho dt$ and $\alpha > 0$, σ_r, $\delta > 0$, $\overline{\gamma} > 0$, σ_γ, and ρ are constants. In addition, define the constant market prices of risk associated with dz_r and dz_γ to be θ_r and θ_γ. Rewrite this model using the affine model notation used in this chapter and solve for the equilibrium price of a zero-coupon bond, $P(t, T)$.

2. Consider the following one-factor quadratic Gaussian model. The single state variable, $x(t)$, follows the risk-neutral process

$$dx(t) = \kappa \left[\overline{x} - x(t)\right] dt + \sigma_x d\widehat{z}$$

and the instantaneous-maturity interest rate is given by $r(t, x) = \alpha + \beta x(t) + \gamma x(t)^2$. Assume κ, $\overline{x}$, α, and γ are positive constants and that $\alpha - \frac{1}{4}\beta^2/\gamma \geq 0$, where β also is a constant. Solve for the equilibrium price of a zero-coupon bond, $P(t, T)$.

3. Show that for the extended Vasicek model when $\overline{r}(t) \equiv \frac{1}{\alpha}\partial f(0, t)/\partial t + f(0, t) + \sigma_r^2 \left(1 - e^{-2\alpha t}\right) / \left(2\alpha^2\right)$, then $P(0, T) = \widehat{E}\left[\exp\left(-\int_0^T r(s)\, ds\right)\right] = \exp\left(-\int_0^T f(0, s)\, ds\right)$.

4. Determine the value of an n-payment interest rate floor using the LIBOR market model.

Chapter 18

Models of Default Risk

The bond pricing models in previous chapters assumed that bonds' promised cashflows are paid with certainty. Therefore, these models are most applicable to valuing default-free bonds issued by a federal government, which would include Treasury bills, notes, and bonds.[1] However, many debt instruments, including corporate bonds, municipal bonds, and bank loans, have default or "credit" risk. Valuing defaultable debt requires an extended modeling approach. We now consider the two primary methods for modeling default risk. The first, suggested in the seminal option pricing paper of Fischer Black and Myron Scholes (Black and Scholes 1973) and developed by Robert Merton (Merton 1974), Francis Longstaff and Eduardo Schwartz (Longstaff and Schwartz 1995), and others is called the "structural" approach. This method values a firm's debt as an explicit function of the value of the firm's assets and its capital structure.

The second "reduced-form" approach more simply assumes that default is a Poisson process with a possibly time-varying default intensity and default

[1]Default can be avoided on government bonds that promise a nominal (currency-valued) payment if the government (or its central bank) has the power to print currency. However, if a federal government relinquishes this power, as is the case for countries that adopted the Euro supplied by the European Central Bank, default on government debt becomes a possibility.

recovery rate. This method views the exogenously specified default process as the reduced form of a more complicated and complex model of a firm's assets and capital structure. Examples of this approach include work by Robert Jarrow, David Lando, and Stuart Turnbull (Jarrow, Lando, and Turnbull 1997), Dilip Madan and Haluk Unal (Madan and Unal 1998), and Darrell Duffie and Kenneth Singleton (Duffie and Singleton 1999). This chapter provides an introduction to the main features of these two methods for incorporating default risk in bond values.

18.1 The Structural Approach

This section focuses on a model similar to that of Robert Merton (Merton 1974).

It specifies the assets, debt, and shareholders' equity of a particular firm. Let $A(t)$ denote the date t value of a firm's assets. The firm is assumed to have a very simple capital structure. In addition to shareholders' equity, it has issued a single zero-coupon bond that promises to pay an amount B at date $T > t$. Also let $\tau \equiv T - t$ be the time until this debt matures. The firm is assumed to pay dividends to its shareholders at the continuous rate $\delta A(t)dt$, where δ is the firm's constant proportion of assets paid in dividends per unit time. The value of the firm's assets is assumed to follow the process

$$dA/A = (\mu - \delta)\, dt + \sigma dz \tag{18.1}$$

where μ denotes the instantaneous expected rate of return on the firm's assets and σ is the constant standard deviation of return on firm assets. Now let $D(t,T)$ be the date t market value of the firm's debt that is promised the payment of B at date T. It is assumed that when the debt matures, the firm pays the promised amount to the debtholders if there is sufficient asset value to do so. If not, the firm defaults (bankruptcy occurs) and the debtholders take ownership of all of the firm's assets. Hence, the payoff to debtholders at date T can be written as

$$\begin{aligned} D(T,T) &= \min[B, A(T)] \\ &= B - \max[0, B - A(T)] \end{aligned} \tag{18.2}$$

From the second line in equation (18.2), we see that the payoff to the debtholders equals the promised payment, B, less the payoff on a European put option written on the firm's assets and having exercise price equal to B. Hence, if we make the usual "frictionless" market assumptions, then the current market value of the debt can be derived to equal the present value of the promised payment less the value of a put option on the dividend-paying assets.[2] If we let $P(t,T)$ be the current date t price of a default-free, zero-coupon bond that pays \$1 at date T and assume that the default-free term structure satisfies the Vasicek model as specified earlier in (9.41) to (9.43), then using Chapter 9's results on the pricing of options when interest rates are random, we obtain

$$\begin{aligned} D(t,T) &= P(t,T)B - P(t,T)BN(-h_2) + e^{-\delta\tau}AN(-h_1) \\ &= P(t,T)BN(h_2) + e^{-\delta\tau}AN(-h_1) \end{aligned} \tag{18.3}$$

where $h_1 = \left[\ln\left[e^{-\delta\tau}A/\left(P(t,T)B\right)\right] + \frac{1}{2}v^2\right]/v$, $h_2 = h_1 - v$, and $v(\tau)$ is given in (9.61). Note that if the default-free term structure is assumed to be deterministic, then we have the usual Black-Scholes value of $v = \sigma\sqrt{\tau}$. The promised yield to maturity on the firm's debt, denoted $R(t,T)$, can be calculated from (18.3) as $R(t,T) = \frac{1}{\tau}\ln[B/D(t,T)]$. Also, its credit spread, which is defined

[2] One needs to assume that the risk of the firm's assets, as determined by the dz process, is a tradeable risk, so that a Black-Scholes hedge involving the firm's debt can be constructed.

as the bond's yield less that of an equivalent maturity default-free bond, can be computed as $R(t,T) - \frac{1}{\tau}\ln[1/P(t,T)]$.

Based on this result, one can also solve for the market value of the firm's shareholder's equity, which we denote as $E(t)$. In the absence of taxes and other transactions costs, the value of investors' claims on the firm's assets, $D(t,T) + E(t)$, must equal the total value of the firm's assets, $A(t)$. This allows us to write

$$\begin{aligned} E(t) &= A(t) - D(t,T) \\ &= A - P(t,T)BN(h_2) - e^{-\delta\tau}AN(-h_1) \\ &= A\left[1 - e^{-\delta\tau}N(-h_1)\right] - P(t,T)BN(h_2) \end{aligned} \tag{18.4}$$

Shareholders' equity is similar to a call option on the firm's assets in the sense that at the debt's maturity date, equity holders receive the payment $\max[A(T) - B, 0]$. Shareholders' limited liability gives them the option of receiving the firm's residual value when it is positive. However, shareholders' equity differs from the standard European call option if the firm pays dividends prior to the debt's maturity. As is reflected in the first term in the last line of (18.4), the firm's shareholders, unlike the holders of standard options, receive these dividends.

Robert Merton (Merton 1974); Chapter 12 in (Merton 1992) gives an in-depth analysis of the comparative statics properties of the debt and equity formulas similar to equations (18.3) and (18.4), as well as the firm's credit spread. Note that an equity formula such as (18.4) can be useful because for firms that have publicly traded shareholders' equity, observation of the firm's market value of equity and its volatility can be used to infer the market value and volatility of the firm's assets. The market value and volatility of the firm's assets can then be used as inputs into (18.3) so that the firm's default-risky debt can be valued. Such an exercise based on the Merton model has been done by the credit-rating firm Moody's KMV to forecast corporate defaults.[3]

The Merton model's assumption that the firm has a single issue of zero-coupon debt is unrealistic, since it is commonly the case that firms have multiple coupon-paying debt issues with different maturities and different seniorities in the event of default. Modeling multiple debt issues and determining the point at which an asset deficiency triggers default is a complex task.[4] In response, some research has taken a different tack by assuming that when the firm's assets hit a lower boundary, default is triggered. This default boundary is presumed to bear a monotonic relation to the firm's total outstanding debt. With the initial value of the firm's assets exceeding this boundary, determining future default amounts to computing the first passage time of the assets through this boundary.

Francis Longstaff and Eduardo Schwartz (Longstaff and Schwartz 1995) developed such a model following the earlier work of Fischer Black and John Cox (Black and Cox 1976). They assume a default boundary that is constant over time and, when assets sink to the level of this boundary, bondholders are assumed to recover an exogenously given proportion of their bonds' face values. This contrasts with the Merton model, where in the case of default, bondholders recover $A(T)$, the stochastic value of firm assets at the bond's maturity date, which results in a loss of $B - A(T)$. In the Longstaff-Schwartz model, possible default occurs at a stochastic date, say, τ, defined by the first (passage) time that $A(\tau) = k$, where k is the predetermined default boundary. Bondholders

[3] For a description of the KMV application of the Merton model for forecasting defaults, see (Crosbie and Bohn 2002). Alan Marcus and Israel Shaked (Marcus and Shaked 1984) apply the Merton model to analyzing the default risk of commercial banks that have publicly traded shareholders' equity.

[4] A study by Edward Jones, Scott Mason, and Eric Rosenfeld (Jones, Mason, and Rosenfeld 1984) is an example.

are assumed to recover $\delta P(\tau, T) B$, where $\delta < 1$ is the recovery rate equaling a proportion of the market value of an otherwise equivalent default-free bond, $P(\tau, T) B$.[5] This exogenous recovery rate, δ, is permitted to differ for bonds with different maturity and seniority characteristics and might be estimated from the historical recovery rates of different types of bonds.

Pierre Collin-Dufresne and Robert Goldstein (Collin-Dufresne and Goldstein 2001) modify the Longstaff-Schwartz model to permit a firm's default boundary to be stochastic. Motivated by the tendency of firms to target their leverage ratios by partially adjusting their debt and equity over time, Collin-Dufresne and Goldstein permit the ratio of firm assets to firm debt (the default boundary) to follow a mean-reverting process with default triggered when this ratio declines to unity.[6] Chunsheng Zhou (Zhou 2001) and Jing-zhi Huang and Ming Huang (Huang and Huang 2003) extend the Longstaff-Schwartz model in another direction by allowing the firm's assets to follow a mixed jump-diffusion process. In this case, assets can suddenly plunge below the default boundary, making default more abrupt than when assets have continuous sample paths. While these "first passage time" models seek to provide more realism than the more simple Merton model, they come at the cost of requiring numerical, rather than closed-form, solutions.[7]

For firms with complicated debt structures, these first passage time models simplify the determination of default by assuming it occurs when a firm's assets

[5] $P(\tau, T) B$ is the market value of a zero-coupon bond paying the face value of B at date T. However, Longstaff and Schwartz do not limit their analysis to defaultable zero-coupon bonds. Indeed, they value both fixed- and floating-coupon bonds assuming a Vasicek model of the term structure. Hence, in general, recovery equals a fixed proportion, δ, of the market value of an otherwise equivalent default-free (fixed- or floating-rate) bond.

[6] More precisely, they assume that the risk-neutral process for the log of the ratio of firm debt to assets, say, $l(t) = \ln[k(t)/A(t)]$, follows an Ornstein-Uhlenbeck process. For an example of a model displaying mean-reverting leverage in the context of commercial bank defaults, see (Pennacchi 2005).

[7] An exception is the closed-form solutions obtained by Stijn Claessens and George Pennacchi (Claessens and Pennacchi 1996), who model default-risky sovereign debt such as Brady bonds.

sink to a specified boundary. The interaction between default and the level and timing of particular promised bond payments are not directly modeled, except as they might affect the specification of the default boundary. In the next section, we consider the reduced-form approach, which goes a step further by not directly modeling either the firm's assets or its overall debt level.

18.2 The Reduced-Form Approach

With the reduced-form method, default need not be tied directly to the dynamics of a firm's assets and liabilities. As a result, this approach provides less insight regarding the link between a firm's balance sheet and its likelihood of default. However, because reduced-form models generate default based on an exogenous Poisson process, they may better capture the effects on default of additional unobserved factors and provide richer dynamics for the term structure of credit spreads.[8] Reduced-form modeling also can be convenient because, as will be shown, defaultable bonds are valued using techniques similar to those used to value default-free bonds.

To illustrate reduced-form modeling, we begin by analyzing a defaultable zero-coupon bond and, later, generalize the results to multiple-payment (coupon) bonds. As in the previous section, let $D(t, T)$ be the date t value of a default-

[8] In most structural models, (Zhou 2001) and (Huang and Huang 2003) are notable exceptions, a firm's assets are assumed to follow a diffusion process that has a continuous sample path. An implication of this is that default becomes highly unlikely for short horizons if the firm currently has a substantial difference between assets and liabilities. Hence, these models generate very small credit spreads for the short-maturity debt of creditworthy corporations, counter to empirical evidence that finds more significant spreads. Small spreads occur because default over a short horizon cannot come as a sudden surprise. This is not the case with reduced-form models, where sudden default is always possible due to its Poisson nature. Hence, these models can more easily match the significant credit spreads on short-term corporate debt. Darrell Duffie and David Lando (Duffie and Lando 2001) present a structural model where investors have less (accounting) information regarding the value of a firm's assets than do the firm's insiders. Hence, like the jump-diffusion models (Zhou 2001); and (Huang and Huang 2003), investors' valuation of the firm's assets can take discrete jumps when inside information is revealed. This model generates a Poisson default intensity equivalent to a particular reduced-form model.

risky, zero-coupon bond that promises to pay B at its maturity date of T. However, unlike the previous section's structural models where default was directly linked to the dynamics of the firm's capital structure, here we assume that a possible default event depends on a reduced-form process that only indirectly may be interpreted as depending on the firm's capital structure and possibly other macroeconomic factors that influence default. Specifically, default for a particular firm's bond is modeled as a Poisson process with a time-varying default intensity. Conditional on default having not occurred prior to date t, the instantaneous probability of default during the interval $(t, t + dt)$ is denoted $\lambda(t)\,dt$, where $\lambda(t)$ is the physical default intensity, or "hazard rate," and is assumed to be nonnegative.[9] The time-varying nature of $\lambda(t)$ may be linked to variation in state variables, as will be shown shortly.

Note from the definition of the instantaneous default intensity, $\lambda(t)$, one can compute the physical probability that the bond does *not* default over the discrete time interval from dates t to τ, where $t < \tau \leq T$. This probability is referred to as the bond's (physical) survival probability over the interval from dates t to τ and is given by

$$E_t\left[e^{-\int_t^\tau \lambda(u)du}\right] \tag{18.5}$$

18.2.1 A Zero-Recovery Bond

To determine $D(t,T)$, an assumption must be made regarding the payoff received by bondholders should the bond default. We begin by assuming that

[9] Recall that in Chapter 11 we modeled jumps in asset prices as following a Poisson process with jump intensity λ. Here, a one-time default follows a Poisson process, and its intensity is explicitly time varying.

bondholders recover nothing if the bond defaults and, later, we generalize this assumption to permit a possible nonzero recovery value. With zero recovery, the bondholders' date T payoff is either $D(T,T) = B$ if there is no default or $D(T,T) = 0$ if default has occurred over the interval from t to T. Applying risk-neutral pricing, the date t value of the zero-recovery bond, denoted $D_Z(t,T)$, can be written as

$$D_Z(t,T) = \widehat{E}_t\left[e^{-\int_t^T r(u)du} D(T,T)\right] \tag{18.6}$$

where $r(t)$ is the date t instantaneous default-free interest rate, and $\widehat{E}_t[\cdot]$ is the date t risk-neutral expectations operator. To compute this expression, we need to determine the expression for $D(T,T)$ in terms of the risk-neutral default intensity, rather than the physical default intensity. The risk-neutral default intensity will account for the market price of risk associated with the Poisson arrival of a default event.

To understand the role of default risk, suppose that both the default-free term structure and the firm's default intensity depend on a set of n state variables, x_i, $i = 1, ..., n$, that follow the multivariate Markov diffusion process[10]

$$d\mathbf{x} = \mathbf{a}(t,\mathbf{x})\,dt + \mathbf{b}(t,\mathbf{x})\,\mathbf{dz} \tag{18.7}$$

where $\mathbf{x} = (x_1...x_n)'$, $\mathbf{a}(t,\mathbf{x})$ is an $n \times 1$ vector, $\mathbf{b}(t,\mathbf{x})$ is an $n \times n$ matrix, and $\mathbf{dz} = (dz_1...dz_n)'$ is an $n \times 1$ vector of independent Brownian motion processes so that $dz_i dz_j = 0$ for $i \neq j$. As in the previous chapter, $\mathbf{x}(t)$ includes macroeconomic factors that affect the default-free term structure, but it now also includes firm-specific factors that affect the likelihood of default for the partic-

[10] For concreteness our presentation assumes an equilibrium Markov state variable environment. However, much of our results on reduced-form pricing of defaultable bonds carries over to a non-Markov, no-arbitrage context, such as the Heath-Jarrow-Morton framework. See (Duffie and Singleton 1999), (Fan and Ritchken 2001), and (Ritchken and Sun 2003).

ular firm. Similar to (17.8), the stochastic discount factor for pricing the firm's default-risky bond will be of the form

$$dM/M = -r(t, \mathbf{x})\,dt - \boldsymbol{\Theta}(t, \mathbf{x})' \mathbf{dz} - \boldsymbol{\psi}(t, \mathbf{x})\,[dq - \lambda(t, \mathbf{x})\,dt] \tag{18.8}$$

where $\boldsymbol{\Theta}(t, \mathbf{x})$ is an $n \times 1$ vector of the market prices of risk associated with the elements of $\mathbf{dz}$ and $\boldsymbol{\psi}(t, \mathbf{x})$ is the market price of risk associated with the actual default event. This default event is recorded by dq, which is a Poisson counting process similar to that described in equation (11.2) of Chapter 11. When default occurs, this Poisson counting process $q(t)$ jumps from 0 (the no-default state) to 1 (the absorbing default state) at which time $dq = 1$.[11] The risk-neutral default intensity, $\widehat{\lambda}(t, \mathbf{x})$, is then given by $\widehat{\lambda}(t, \mathbf{x}) = [1 - \boldsymbol{\psi}(t, \mathbf{x})]\,\lambda(t, \mathbf{x})$. Note that in this modeling context, default is a "doubly stochastic" process, also referred to as a Cox process.[12] Default depends on the Brownian motion vector $\mathbf{dz}$ that drives $\mathbf{x}$ and determines how the likelihood of default, $\widehat{\lambda}(t, \mathbf{x})$, changes over time, but it also depends on the Poisson process dq that determines the arrival of default. Hence, default risk reflects two types of risk premia, $\boldsymbol{\Theta}(t, \mathbf{x})$ and $\boldsymbol{\psi}(t, \mathbf{x})$.

Based on the calculation of survival probability in (18.5), the value of the zero-recovery defaultable bond is

$$D_Z(t, T) = \widehat{E}_t\left[e^{-\int_t^T r(u)du} e^{-\int_t^T \widehat{\lambda}(u)du} B\right] = \widehat{E}_t\left[e^{-\int_t^T \left[r(u)+\widehat{\lambda}(u)\right]du}\right] B \tag{18.9}$$

Equation (18.9) shows that valuing this zero-recovery defaultable bond is similar

[11] Recall from the discussion in Chapter 11 that jumps in an asset's value, as would occur when a bond defaults, cannot always be hedged. Thus, in general, it may not be possible to determine $\boldsymbol{\psi}(t, \mathbf{x})$ based on a no-arbitrage restriction. This market price of default risk may need to be determined from an equilibrium model of investor preferences.

[12] Named after the statistician Sir David Cox (Cox 1955).

to valuing a default-free bond except that we use the discount rate of $r(u)+\widehat{\lambda}(u)$ rather than just $r(u)$. Given specific functional forms for $r(t,\mathbf{x})$, $\widehat{\lambda}(t,\mathbf{x})$, and the risk-neutral state variable process (specifications of (18.7) and $\boldsymbol{\Theta}(t,\mathbf{x})$), the expression in (18.9) can be computed.

18.2.2 Specifying Recovery Values

The value of a bond that has a possibly nonnegative recovery value in the event of default equals the value in (18.9) plus the present value of the amount recovered in default. Suppose that if the bond defaults at date τ where $t < \tau \leq T$, bondholders recover an amount $w(\tau,\mathbf{x})$ at date τ. Now note that the risk-neutral probability density of defaulting at time τ is

$$e^{-\int_t^\tau \widehat{\lambda}(u)du}\widehat{\lambda}(\tau) \tag{18.10}$$

In (18.10), $\widehat{\lambda}(\tau)$ is discounted by $\exp\left[-\int_t^\tau \widehat{\lambda}(u)\,du\right]$ because default at date τ is conditioned on not having defaulted previously. Therefore, the present value of recovery in the event of default, $D_R(t,T)$, is computed by integrating the expected discounted value of recovery over all possible default dates from t to T:

$$\begin{aligned} D_R(t,T) &= \widehat{E}_t\left[\int_t^T e^{-\int_t^\tau r(u)du} w(\tau)\, e^{-\int_t^\tau \widehat{\lambda}(u)du}\widehat{\lambda}(\tau)\,d\tau\right] \\ &= \widehat{E}_t\left[\int_t^T e^{-\int_t^\tau \left[r(u)+\widehat{\lambda}(u)\right]du}\widehat{\lambda}(\tau)\,w(\tau)\,d\tau\right] \end{aligned} \tag{18.11}$$

Putting this together with (18.9) gives the bond's total value, $D(t,T) = D_Z(t,T) + D_R(t,T)$, as

$$D(t,T) = \widehat{E}_t \left[e^{-\int_t^T [r(s)+\widehat{\lambda}(s)]ds} B + \int_t^T e^{-\int_t^\tau [r(s)+\widehat{\lambda}(s)]ds} \widehat{\lambda}(\tau) w(\tau) d\tau \right] \tag{18.12}$$

Recovery Proportional to Par Value

Let us consider some particular specifications for $w(\tau, \mathbf{x})$. One assumption used by several researchers is that bondholders recover at the default date τ a proportion of the bond's face, or par, value; that is, $w(\tau, \mathbf{x}) = \delta(\tau, \mathbf{x}) B$, where $\delta(\tau, \mathbf{x})$ is usually assumed to be a constant, say, $\overline{\delta}$.[13] In this case, (18.11) can be written as

$$D_R(t,T) = \overline{\delta} B \int_t^T k(t,\tau) d\tau \tag{18.13}$$

where

$$k(t,\tau) \equiv \widehat{E}_t \left[e^{-\int_t^\tau [r(u)+\widehat{\lambda}(u)]du} \widehat{\lambda}(\tau) \right] \tag{18.14}$$

has a closed-form solution when $r(u, \mathbf{x})$ and $\widehat{\lambda}(u, \mathbf{x})$ are affine functions of $\mathbf{x}$ and the vector $\mathbf{x}$ in (18.7) has a risk-neutral process that is also affine.[14] In this case, the recovery value in (18.13) can be computed by numerical integration of $k(t,\tau)$ over the interval from t to T.

Recovery Proportional to Par Value, Payable at Maturity

An alternative recovery assumption is that if default occurs at date τ, the bondholders recover a proportion $\delta(\tau, \mathbf{x})$ of the bond's face value, B, payable at the

[13] Work by Darrell Duffie (Duffie 1998), David Lando (Lando 1998), and Dilip Madan and Haluk Unal (Madan and Unal 1998) makes this assumption. As reported by Gregory Duffee (Duffee 1999), the recovery rate, $\overline{\delta}$, estimated by Moody's for senior unsecured bondholders, is approxmately 44 percent.

[14] This is shown in (Duffie, Pan, and Singleton 2000).

maturity date T.[15] This is equivalent to assuming that the bondholders recover a proportion $\delta(\tau, \mathbf{x})$ of the market value of a default-free discount bond paying B at date T; that is, $w(\tau, \mathbf{x}) = \delta(\tau, \mathbf{x}) P(\tau, T) B$. Under this assumption, (18.11) becomes

$$\begin{aligned} D_R(t,T) &= \widehat{E}_t\left[\int_t^T e^{-\int_t^\tau \left[r(u)+\widehat{\lambda}(u)\right]du}\widehat{\lambda}(\tau)\,\delta(\tau,\mathbf{x})\,e^{-\int_\tau^T r(u)du}Bd\tau\right] \\ &= \widehat{E}_t\left[\int_t^T e^{-\int_t^\tau \widehat{\lambda}(u)du}\widehat{\lambda}(\tau)\,\delta(\tau,\mathbf{x})\,e^{-\int_t^T r(u)du}Bd\tau\right] \\ &= \widehat{E}_t\left[e^{-\int_t^T r(u)du}\int_t^T e^{-\int_t^\tau \widehat{\lambda}(u)du}\widehat{\lambda}(\tau)\,\delta(\tau,\mathbf{x})\,d\tau\right]B \qquad (18.15) \end{aligned}$$

For the specific case of $\delta(\tau, x) = \overline{\delta}$, a constant, this expression can be simplified by noting that the term $\int_t^T \exp\left[-\int_t^\tau \widehat{\lambda}(u)\,du\right]\widehat{\lambda}(\tau)\,d\tau$ is the total risk-neutral probability of default for the period from date t to the maturity date T. Therefore, it must equal $1 - \exp\left[-\int_t^T \widehat{\lambda}(u)\,du\right]$; that is, 1 minus the probability of surviving over the same period. Making this substitution and using (18.9), we have

$$\begin{aligned} D_R(t,T) &= \widehat{E}_t\left[e^{-\int_t^T r(u)du}\left(1 - e^{-\int_t^T \widehat{\lambda}(u)du}\right)\right]\delta B \\ &= \widehat{E}_t\left[e^{-\int_t^T r(u)du} - e^{-\int_t^T \left[r(u)+\widehat{\lambda}(u)\right]du}\right]\delta B \\ &= \overline{\delta}BP(t,T) - \overline{\delta}D_Z(t,T) \qquad (18.16) \end{aligned}$$

Therefore, the total value of the bond is

$$D(t,T) = D_Z(t,T) + D_R(t,T) = \left(1-\overline{\delta}\right)D_Z(t,T) + \overline{\delta}BP(t,T) \qquad (18.17)$$

[15] This specification has been studied by Robert Jarrow and Stuart Turnbull (Jarrow and Turnbull 1995) and David Lando (Lando 1998).

Hence, this recovery assumption amounts to requiring only a solution for the value of a zero-recovery bond.

Recovery Proportional to Market Value

Let us consider one additional recovery assumption analyzed by Darrell Duffie and Kenneth Singleton (Duffie and Singleton 1999). When default occurs, bondholders are assumed to recover a proportion of what was the bond's market value just prior to default. This is equivalent to assuming that the bond's market value jumps downward at the default date τ, suffering a proportional loss of $L(\tau, \mathbf{x})$. Specifically, at default $D(\tau^-, T)$ jumps to

$$D\left(\tau^+, T\right) = w(\tau, \mathbf{x}) = D\left(\tau^-, T\right)\left[1 - L(\tau, \mathbf{x})\right] \tag{18.18}$$

By specifying a proportional loss in value at the time of default, the bond's dynamics become similar to the jump-diffusion model of asset prices presented in Chapter 11. Treating the defaultable bond as a contingent claim and applying Itô's lemma, its process prior to default is similar to equation (11.6):

$$dD(t,T)/D(t,T) = (\alpha_D - \lambda k_D)\,dt + \boldsymbol{\sigma}'_D \mathbf{dz} - L(t, \mathbf{x})\,dq \tag{18.19}$$

where α_D and the $n \times 1$ vector $\boldsymbol{\sigma}_D$ are given by the usual Itô's lemma expressions similar to (11.7) and (11.8). From (11.3) and (18.18), we have that when a jump occurs $\left[D(\tau^+, T) - D(\tau^-, T)\right]/D(\tau^-, T) = -L(\tau, \mathbf{x})$, which verifies the term $-L(t, \mathbf{x})\,dq$. Also, from (11.10), k_D, the expected jump size, is given by $k_D(\tau^-) \equiv E_{\tau^-}\left[D(\tau^+, T) - D(\tau^-, T)\right]/D(\tau^-, T) = -L(\tau, \mathbf{x})$, so that the drift term in (18.19) becomes $\alpha_D + \lambda(t, \mathbf{x})\,L(t, \mathbf{x})$.

Now under the risk-neutral measure, the defaultable bond's total expected rate of return, α_D, must equal the instantaneous-maturity, default-free rate,

$r(t)$. Thus, we can write the bond's risk-neutral process prior to default as

$$dD(t,T)/D(t,T) = \left(r(t,\mathbf{x}) + \widehat{\lambda}(t,\mathbf{x})\widehat{L}(t,\mathbf{x})\right)dt + \boldsymbol{\sigma}_D'\mathbf{d}\widehat{\mathbf{z}} - \widehat{L}(t,\mathbf{x})\,dq \quad (18.20)$$

where $\widehat{L}(t,\mathbf{x})$ is the risk-neutral expected proportional loss given default.[16] The intuition of (18.20) is that because the bond has a risk-neutral expected loss given default of $\widehat{L}(t,\mathbf{x})$, and the risk-neutral instantaneous probability of default ($dq = 1$) is $\widehat{\lambda}(t,\mathbf{x})$, when the bond does not default it must earn an excess expected return of $\widehat{\lambda}(t,\mathbf{x})\widehat{L}(t,\mathbf{x})$ to make its unconditional risk-neutral expected return equal $r(t)$. Based on a derivation similar to that used to obtain (11.17) and (17.6), one can show that the defaultable bond's value satisfies the equilibrium partial differential equation

$$\tfrac{1}{2}\mathrm{Trace}\left[\mathbf{b}(t,\mathbf{x})\,\mathbf{b}(t,\mathbf{x})'\,\mathbf{D}_{xx}\right] + \widehat{\mathbf{a}}(t,\mathbf{x})'\,\mathbf{D}_x - R(t,\mathbf{x})\,D + D_t = 0 \qquad (18.21)$$

where $\mathbf{D}_x$ denotes the $n \times 1$ vector of first derivatives of $D(t,\mathbf{x})$ with respect to each of the factors and, similarly, $\mathbf{D}_{xx}$ is the $n \times n$ matrix of second-order mixed partial derivatives. In addition, $\widehat{\mathbf{a}}(t,\mathbf{x}) = \mathbf{a}(t,\mathbf{x}) - \mathbf{b}(t,\mathbf{x})\,\boldsymbol{\Theta}$ is the risk-neutral drift of the factor process (18.7), and $R(t,\mathbf{x}) \equiv r(t,\mathbf{x}) + \widehat{\lambda}(t,\mathbf{x})\widehat{L}(t,\mathbf{x})$ is the defaultable bond's risk-neutral drift in the process (18.20). Note that if the bond reaches the maturity date, T, without defaulting, then $D(T,T) = B$. This is the boundary condition for (18.21). The PDE (18.21) is in the form of a PDE for a standard contingent claim except that $R(t,\mathbf{x})$ has replaced $r(t,\mathbf{x})$

[16] As with the risk-neutral default intensity, $\widehat{\lambda}(t,\mathbf{x})$, there may be a market price of recovery risk associated with $\widehat{L}(t,\mathbf{x})$ that distinguishes it from the physical expected loss at default, $L(t,\mathbf{x})$. This market price of recovery risk cannot, in general, be determined from a no-arbitrage restriction because recovery risk may be unhedgeable. Most commonly, modelers simply posit functional forms for risk-neutral variables in order to derive formulas for defaultable bond values. Differences between risk-neutral default intensities and losses at default and their physical counterparts might be inferred based on the market prices of defaultable bonds and historical (physical) default and recovery rates.

in the standard PDE. This insight allows us to write the PDE's Feynman-Kac solution as[17]

$$D(t,T) = \widehat{E}_t\left[e^{-\int_t^T R(u,\mathbf{x})du}\right]B \tag{18.22}$$

where $R(t,\mathbf{x}) \equiv r(t,\mathbf{x}) + \widehat{\lambda}(t,\mathbf{x})\widehat{L}(t,\mathbf{x})$ can be viewed as the "default-adjusted" discount rate. The product $s(t,\mathbf{x}) \equiv \widehat{\lambda}(t,\mathbf{x})\widehat{L}(t,\mathbf{x})$ is interpreted as the "credit spread" on an instantaneous-maturity, defaultable bond. Since $\widehat{\lambda}(t,\mathbf{x})$ and $\widehat{L}(t,\mathbf{x})$ are not individually identified in (18.22), when implementing this formula, one can simply specify a single functional form for $s(t,\mathbf{x})$.

18.2.3 Examples

Because default intensities and/or credit spreads must be nonnegative, a popular stochastic process for modeling these variables is the mean-reverting, square root process used in the term structure model of John Cox, Jonathan Ingersoll, and Stephen Ross (Cox, Ingersoll, and Ross 1985b). To take a very simple example, suppose that $\mathbf{x} = (x_1\ x_2)'$ is a two-dimensional vector, $\widehat{\mathbf{a}}(t,\mathbf{x}) = (\kappa_1(\overline{x}_1 - x_1)\quad \kappa_2(\overline{x}_2 - x_2))'$, and $\mathbf{b}(t,\mathbf{x})$ is a diagonal matrix with first and second diagonal elements of $\sigma_1\sqrt{x_1}$ and $\sigma_2\sqrt{x_2}$, respectively. If one assumes $r(t,\mathbf{x}) = x_1(t)$ and $\widehat{\lambda}(t,\mathbf{x}) = x_2(t)$, this has the implication that the default-free term structure and the risk-neutral default intensity are independent. Arguably, this is unrealistic since empirical work tends to find a negative correlation between default-free interest rates and the likelihood of corporate defaults.[18] Allowing for nonzero correlation between $r(t,\mathbf{x})$ and $\widehat{\lambda}(t,\mathbf{x})$ while restricting each to be positive is certainly feasible but comes at the cost of requiring numer-

[17] Recall from Chapter 10 that (10.17) was shown to be the Feynman-Kac solution to the Black-Scholes PDE (10.7). See Darrell Duffie and Kenneth Singleton (Duffie and Singleton 1999) for an alternative derivation of (18.22) that does not involve specification of factors or the bond's PDE.

[18] This evidence is presented in work by Gregory Duffee (Duffee 1999) and Pierre Collin-Dufresne and Bruno Solnik (Collin-Dufresne and Solnik 2001).

ical, rather than closed-form, solutions for defaultable bond values.[19] Hence, for simplicity of presentation, we maintain the independence assumption in the examples that follow.

With $r(t, \mathbf{x}) = x_1(t)$ and denoting $\overline{x}_1 = \overline{r}$, we obtain the Cox, Ingersoll, and Ross formula for the value of a default-free discount bond:[20]

$$P(t,T) = A_1(\tau)\, e^{-B_1(\tau) r(t)} \tag{18.23}$$

where

$$A_1(\tau) \equiv \left[\frac{2\theta_1 e^{(\theta_1+\kappa_1)\frac{\tau}{2}}}{(\theta_1+\kappa_1)(e^{\theta_1\tau}-1)+2\theta_1}\right]^{2\kappa_1\overline{r}/\sigma_1^2} \tag{18.24}$$

$$B_1(\tau) \equiv \frac{2\left(e^{\theta_1\tau}-1\right)}{(\theta_1+\kappa_1)(e^{\theta_1\tau}-1)+2\theta_1} \tag{18.25}$$

and $\theta_1 \equiv \sqrt{\kappa_1^2 + 2\sigma_1^2}$. Also with $\widehat{\lambda}(t, \mathbf{x}) = x_2(t)$ and denoting $\overline{x}_2 = \overline{\lambda}$, then based on (18.9) and the assumed independence of $r(t)$ and $\widehat{\lambda}(t)$, we can write the value of the zero-recovery bond as

$$\begin{aligned} D_Z(t,T) &= \widehat{E}_t\left[e^{-\int_t^T \left[r(s)+\widehat{\lambda}(s)\right]ds}\right]B \\ &= \widehat{E}_t\left[e^{-\int_t^T r(s)ds}\right]\widehat{E}_t\left[e^{-\int_t^T \widehat{\lambda}(s)ds}\right]B \\ &= P(t,T)\,V(t,T)\,B \end{aligned} \tag{18.26}$$

[19] For models with more flexible correlation structures that require numerical solutions, see examples given by Darrell Duffie and Kenneth Singleton (Duffie and Singleton 1999). Some research has dropped the restriction that $r(t)$ and $\widehat{\lambda}(t)$ (or $s(t) = \widehat{\lambda}(t)\, L(t)$) be positive by assuming these variables follow multivariate affine Gaussian processes. This permits general correlation between default-free interest rates and default intensities as well as closed-form solutions for defaultable bonds. The model in work by C.V.N. Krishnan, Peter Ritchken, and James Thomson (Krishnan, Ritchken, and Thomson 2004) is an example of this.

[20] The formula in (18.23) to (18.25) is the same as (13.51) to (13.53) except that it is written in terms of the parameters of the risk-neutral, rather than physical, process for $r(t)$. Hence, relative to our earlier notation, $\kappa_1 = \kappa + \psi$, where the market price of interest-rate risk equals $\theta(t) = -\psi\sqrt{r}/\sigma_1$.

where

$$V(t,T) = A_2(\tau)\, e^{-B_2(\tau)\widehat{\lambda}(t)} \tag{18.27}$$

and where $A_2(\tau)$ is the same as $A_1(\tau)$ in (18.24), and $B_2(\tau)$ is the same as $B_1(\tau)$ in (18.25) except that κ_2 replaces κ_1, σ_2 replaces σ_1, $\overline{\lambda}$ replaces $\overline{r}$, and $\theta_2 \equiv \sqrt{\kappa_2^2 + 2\sigma_2^2}$ replaces θ_1.

If we assume that recovery is a fixed proportion, $\overline{\delta}$, of par value, payable at maturity, then based on (18.17) the value of the defaultable bond equals

$$\begin{aligned} D(t,T) &= \left(1-\overline{\delta}\right) D_Z(t,T) + \overline{\delta} B P(t,T) \\ &= \left[\overline{\delta} + \left(1-\overline{\delta}\right) V(t,T)\right] P(t,T)\, B \end{aligned} \tag{18.28}$$

In (18.27), $V(t,T)$ is analogous to a bond price in the standard Cox, Ingersoll, and Ross term structure model, and as such it will be inversely related to $\widehat{\lambda}(t)$ and strictly less than 1 whenever $\widehat{\lambda}(t)$ is strictly positive, which can be ensured when $2\kappa_2\overline{\lambda} \geq \sigma_2^2$. Thus, (18.28) confirms that the defaultable bond's value declines as its risk-neutral default intensity rises.

A slightly different defaultable bond formula can be obtained when recovery is assumed to be proportional to market value and $s(t,\mathbf{x}) \equiv \widehat{\lambda}(t,\mathbf{x})\, \widehat{L}(t,\mathbf{x}) = x_2$ with the notation $\overline{x}_2 = \overline{s}$. In this case, (18.22) becomes

$$\begin{aligned} D(t,T) &= \widehat{E}_t\left[e^{-\int_t^T [r(u)+s(u)]du}\right] B \\ &= \widehat{E}_t\left[e^{-\int_t^T r(u)du}\right] \widehat{E}_t\left[e^{-\int_t^T s(u)du}\right] B \\ &= P(t,T)\, S(t,T)\, B \end{aligned} \tag{18.29}$$

where

$$S(t,T) = A_2(\tau)\, e^{-B_2(\tau)s(t)} \tag{18.30}$$

and where $A_2(\tau)$ is the same as $A_1(\tau)$ in (18.24) and $B_2(\tau)$ is the same as $B_1(\tau)$ in (18.25) except that κ_2 replaces κ_1, σ_2 replaces σ_1, $\overline{s}$ replaces $\overline{r}$, and $\theta_2 \equiv \sqrt{\kappa_2^2 + 2\sigma_2^2}$ replaces θ_1. This defaultable bond is priced similarly to a default-free bond except that the instantaneous-maturity interest rate, $R(t) = r(t) + s(t)$, is now the sum of two nonnegative square root processes. Hence, the defaultable bond is inversely related to $s(t)$ and can be strictly less than the default-free bond as $s(t)$ can always be positive when $2\kappa_2\overline{s} \geq \sigma_2^2$.

Coupon Bonds

Valuing the defaultable coupon bond of a particular issuer (e.g., corporation) is straightforward given the preceding analysis of defaultable zero-coupon bonds. Suppose that the issuer's coupon bond promises n cashflows, with the i^{th} promised cashflow being equal to c_i and being paid at date $T_i > t$. Then the value of this coupon bond in terms of our zero-coupon bond formulas is

$$\sum_{i=1}^{n} D(t, T_i) \frac{c_i}{B} \tag{18.31}$$

Credit Default Swaps

Our results can also be applied to valuing credit derivatives. A credit default swap is a popular credit derivative that typically has the following structure. One party, the protection buyer, makes periodic payments until the contract's maturity date as long as a particular issuer, bond, or loan does not default. The other party, the protection seller, receives these payments in return for paying the difference between the bond or loan's par value and its recovery value if default occurs prior to the maturity of the swap contract. At the initial agreement date of this swap contract, the periodic payments are set such that the initial contract has a zero market value.

We can use our previous analysis to value each side of this swap. Let the contract specify equal period payments of c at future dates $t + \Delta$, $t + 2\Delta$, ..., $t + n\Delta$.[21] Then recognizing that these payments are contingent on default not occurring and that they have zero value following a possible default event, their market value equals

$$\frac{c}{B} \sum_{i=1}^{n} D_Z\left(t, t + i\Delta\right) \tag{18.32}$$

where $D_Z\left(t, T\right)$ is the value of the zero-recovery bond given in (18.9). If we let $w\left(\tau, \mathbf{x}\right)$ be the recovery value of the defaultable bond (or loan) underlying the swap contract, then assuming this bond's maturity date is $T \geq t + n\Delta$, the value of the swap protection can be computed similarly to (18.11) as

$$\widehat{E}_t\left[\int_t^{t+n\Delta} e^{-\int_t^\tau \left[r(u)+\widehat{\lambda}(u)\right]du}\widehat{\lambda}\left(\tau\right)\left[B - w\left(\tau\right)\right]d\tau\right] \tag{18.33}$$

The protection seller's payment in the event of default, $B - w\left(\tau\right)$, is often simplified by assuming recovery is a fixed proportion of par value, that is, $B - w\left(\tau\right) = B - \overline{\delta}B = B\left(1 - \overline{\delta}\right)$. For this special case, (18.33) becomes

$$B\left(1 - \overline{\delta}\right)\int_t^{t+n\Delta} k\left(t, \tau\right)d\tau \tag{18.34}$$

where $k\left(t, \tau\right)$ is defined in (18.14). Given assumptions regarding the functional forms of $r\left(t, \mathbf{x}\right)$, $\widehat{\lambda}\left(t, \mathbf{x}\right)$, and $w\left(t, \mathbf{x}\right)$, and the state variables $\mathbf{x}$, the value of the swap payments, c, that equates (18.32) to (18.33) can be determined.

A general issue that arises when implementing the reduced-form approach to valuing risky debt is determining the proper current values $\widehat{\lambda}\left(t\right)$, $s\left(t\right)$, or $w\left(t\right)$ that may not be directly observable. One or more of these default variables might be inferred by setting the actual market prices of one or more of an issuer's

[21] A period of Δ = one-half year is common since these payments often coincide with an underlying coupon bond making semiannual payments.

bonds to their theoretical formulas. Then, based on the "implied" values of $\widehat{\lambda}(t)$, $s(t)$, or $w(t)$, one can determine whether a given bond of the same issuer is over- or underpriced relative to other bonds. Alternatively, these implied default variables could be used to set the price of a new bond of the same issuer or a credit derivative (such as a default swap) written on the issuer's bonds.

18.3 Summary

Research on credit risk has grown rapidly in recent years. In part, the expansion of this literature derives from a greater interest by financial institutions in credit risk management and credit derivatives.[22] New risk management practices and credit derivatives are being spawned as the techniques for quantifying and pricing credit risk evolve. This chapter introduced the two main branches of modeling defaultable fixed-income securities. The structural approach models default based on the interaction between a firm's assets and its liabilities. Potentially, it can improve our understanding between capital structure and corporate bond and loan prices. In contrast, the reduced-form method abstracts from specific characteristics of a firm's financial structure. However, it can permit a more flexible modeling of default probabilities and may better describe actual the prices of an issuer's debt.

While this chapter has been limited to models of corporate defaults, the credit risk literature also encompasses additional topics such as consumer credit risk and the credit risk of (securitized) portfolios of loans and bonds. Inter-

[22] Interest in risk management has been stimulated by the adoption of risk-based capital standards formulated by the Basel Committee on Banking Supervision. This committee is composed of bank supervisors of the major developed countries. International bank capital standards were first devised in 1988 and are referred to as the Basel Capital Accord. A framework for revised capital standards that depend more intricately on credit and other risks, known as Basel II, was issued by the committee in June of 2004. The Basel II rules link a bank's minimum capital to its level of credit risk on bonds, loans, and credit derivatives. See (Basel Committee on Banking Supervision 2005).

est by both academics and practitioners in the broad field of credit risk will undoubtedly continue.

18.4 Exercises

1. Consider the example given in the "structural approach" to modeling default risk. Maintain the assumptions made in the chapter but now suppose that a third party guarantees the firm's debtholders that if the firm defaults, the debtholders will receive their promised payment of B. In other words, this third-party guarantor will make a payment to the debtholders equal to the difference between the promised payment and the firm's assets if default occurs. (Banks often provide such a guarantee in the form of a letter of credit. Insurance companies often provide such a guarantee in the form of bond insurance.)

   What would be the fair value of this bond insurance at the initial date, t? In other words, what is the competitive bond insurance premium charged at date t?

2. Consider a Merton-type "structural" model of credit risk (Merton 1974). A firm is assumed to have shareholders' equity and two zero-coupon bonds that both mature at date T. The first bond is "senior" debt and promises to pay B_1 at maturity date T, while the second bond is "junior" (or subordinated) debt and promises to pay B_2 at maturity date T. Let $A(t)$, $D_1(t)$, and $D_2(t)$ be the date t values of the firm's assets, senior debt, and junior debt, respectively. Then the maturity values of the bonds are

$$D_1(T) = \begin{cases} B_1 & \text{if } A(T) \geq B_1 \\ A(T) & \text{otherwise} \end{cases}$$

$$D_2(T) = \begin{cases} B_2 & \text{if } A(T) \geq B_1 + B_2 \\ A(T) - B_1 & \text{if } B_1 + B_2 > A(T) \geq B_1 \\ 0 & \text{otherwise} \end{cases}$$

The firm is assumed to pay no dividends to its shareholders, and the value of shareholders' equity at date T, $E(T)$, is assumed to be

$$E(T) = \begin{cases} A(T) - (B_1 + B_2) & \text{if } A(T) \geq B_1 + B_2 \\ 0 & \text{otherwise} \end{cases}$$

Assume that the value of the firm's assets follows the process

$$dA/A = \mu dt + \sigma dz$$

where μ denotes the instantaneous expected rate of return on the firm's assets and σ is the constant standard deviation of return on firm assets. In addition, the continuously compounded, risk-free interest rate is assumed to be the constant r. Let the current date be t, and define the time until the debt matures as $\tau \equiv T - t$.

a. Give a formula for the current, date t, value of shareholders' equity, $E(t)$.

b. Give a formula for the current, date t, value of the senior debt, $D_1(t)$.

c. Using the results from parts (a.) and (b.), give a formula for the current, date t, value of the junior debt, $D_2(t)$.

3. Consider a portfolio of m different defaultable bonds (or loans), where the i^{th} bond has a default intensity of $\lambda_i(t, \mathbf{x})$ where $\mathbf{x}$ is a vector of state variables that follows the multivariate diffusion process in (18.7). Assume that the only source of correlation between the bonds' defaults is through

their default intensities. Suppose that the maturity dates for the bonds all exceed date $T > t$. Write down the expression for the probability that none of the bonds in the portfolio defaults over the period from date t to date T.

4. Consider the standard "plain vanilla" swap contract described in Chapter 17. In equation (17.74) it was shown that under the assumption that each party's payments were default free, the equilibrium swap rate agreed to at the initiation of the contract, date T_0, equals

$$s_{0,n}(T_0) = \frac{1 - P(T_0, T_{n+1})}{\tau \sum_{j=1}^{n+1} P(T_0, T_j)}$$

where for this contract, fixed-interest-rate coupon payments are exchanged for floating-interest-rate coupon payments at the dates T_1, T_2, ...,T_{n+1}, where $T_{j+1} = T_j + \tau$ and τ is the maturity of the LIBOR of the floating-rate coupon payments. This swap rate formula is valid when neither of the parties have credit risk. Suppose, instead, that they both have the same credit risk, and it is equivalent to the credit risk reflected in LIBOR interest rates. (Recall that LIBOR reflects the level of default risk for a large international bank.) Moreover, assume a reduced-form model of default with recovery proportional to market value, so that the value of a LIBOR discount bond promising \$1 at maturity date T_j is given by (18.22):

$$D(T_0, T_j) = \widehat{E}_{T_0}\left[e^{-\int_{T_0}^{T_j} R(u,\mathbf{x})du}\right]$$

where the default-adjusted instantaneous discount rate $R(t, \mathbf{x}) \equiv r(t, \mathbf{x}) + \widehat{\lambda}(t, \mathbf{x}) \widehat{L}(t, \mathbf{x})$ is assumed to be the same for both parties. Assume that if default occurs at some date $\tau < T_{n+1}$, the counterparty whose position is in the money (whose position has positive value) suffers a proportional loss of $L(\tau, \mathbf{x})$ in that position. Show that under these assumptions, the equilibrium swap rate is

$$s_{0,n}(T_0) = \frac{1 - D(T_0, T_{n+1})}{\tau \sum_{j=1}^{n+1} D(T_0, T_j)}$$