Part IV

Consumption, Portfolio Choice, and Asset Pricing in Continuous Time

Chapter 12

Continuous-Time Consumption and Portfolio Choice

Until now our applications of continuous-time stochastic processes have focused on the valuation of contingent claims. In this chapter we revisit the topic introduced in Chapter 5, namely, an individual's intertemporal consumption and portfolio choice problem. However, rather than assume a discrete-time setting, we now examine this problem where asset prices are subject to continuous, random changes and an individual can adjust consumption and portfolio allocations at any time. Specifically, this chapter assumes that an individual maximizes a time-separable expected utility function that depends on the rate of consumption at all future dates. The savings of this individual are allocated among assets whose returns follow diffusion processes of the type first introduced in Chapter 8. Hence, in this environment, the values of the individual's portfolio holdings and total wealth change constantly and, in general, it is optimal for the individual to make continuous rebalancing decisions.

The continuous-time consumption and portfolio choice problem just described was formulated and solved in two papers by Robert Merton (Merton 1969); (Merton 1971). This work was the foundation of his model of intertemporal asset pricing (Merton 1973a), which we will study in the next chapter. As we discuss next, allowing individuals to rebalance their portfolios continuously can lead to qualitatively different portfolio choices compared to those where portfolios can only be adjusted at discrete dates. This, in turn, means that the asset pricing implications of individuals' decisions in continuous time can sometimes differ from those of a discrete-time model. Continuous trading may enable markets to be dynamically complete and lead to sharper asset pricing results. For this reason, continuous-time consumption and portfolio choice models are often used in financial research on asset pricing. Much of our analysis in later chapters will be based on such models.

By studying consumption and portfolio choices in continuous time, the effects of time variation in assets' return distributions, that is, changing investment opportunities, become transparent. As will be shown, individuals' portfolio choices include demands for assets that are the same as those derived from the single-period mean-variance analysis of Chapter 2. However, portfolio choices also include demands for assets that hedge against changes in investment opportunities. This is a key insight that differentiates single-period and multiperiod models and has implications for equilibrium asset pricing.

The next section outlines the assumptions of an individual's consumption and portfolio choice problem for a continuous-time environment. Then, similar to what was done in solving for an individual's decisions in discrete time, we introduce and apply a continuous-time version of stochastic dynamic programming to derive consumption and portfolio demands. This technique leads to a nonlinear partial differential equation that can be solved to obtain optimal decision rules. Portfolio behaviors for both constant investment opportunities and changing investment opportunities are analyzed. We next present an alternative martingale approach to finding an individual's optimal consumption and portfolio choices. This martingale technique is most applicable to situations when markets are dynamically complete and involves computing an expectation of future discounted consumption rates or solving a Black-Scholes-type linear partial differential equation for wealth. We illustrate this solution method by an example where an individual faces risky-asset returns that are negatively correlated with investment opportunities.

12.1 Model Assumptions

Let us assume that an individual allocates his wealth between n different risky assets plus a risk-free asset. Define $S_i(t)$ as the price of the i^{th} risky asset at date t. This asset's instantaneous rate of return is assumed to satisfy the process[1]

$$dS_i(t) \,/\, S_i(t) \;=\; \mu_i\,(x,t)\;dt \;+\; \sigma_i\,(x,t)\;dz_i \tag{12.1}$$

where $i = 1, ..., n$ and $(\sigma_i\,dz_i)(\sigma_j\,dz_j) \;= \sigma_{ij}\,dt$. In addition, let the instantaneous risk-free return be denoted as $r\,(x,t)$. It is assumed that μ_i, σ_i, and r may be functions of time and a $k \times 1$ vector of state variables, which we denote by $x\,(t) = (x_1...x_k)'$. When the μ_i, σ_i, and/or r are time varying, the investor is said to face changing investment opportunities. The state variables affecting the moments of the asset prices can, themselves, follow diffusion processes. Let

[1] Equation (12.1) expresses a risky asset's rate of return process in terms of its proportional price change, dS_i/S_i. However, if the asset pays cashflows (e.g., dividends or coupons), then $S_i\,(t)$ can be reinterpreted as the value of an investment in the risky asset where all cashflows are reinvested. What is essential is the asset's return process, not whether returns come in the form of cash payouts or capital gains.

the i^{th} state variable follow the process

$$dx_i = a_i\left(x,t\right)dt + b_i\left(x,t\right)d\zeta_i \tag{12.2}$$

where $i = 1, ..., k$. The process $d\zeta_i$ is a Brownian motion with $(b_i\, d\zeta_i)(b_j\, d\zeta_j) = b_{ij}\, dt$ and $(\sigma_i\, dz_i)(b_j\, d\zeta_j) = \phi_{ij}\, dt$. Hence, equations (12.1) and (12.2) indicate that up to $n + k$ sources of uncertainty (Brownian motion processes) affect the distribution of asset returns.

We denote the value of the individual's wealth portfolio at date t as W_t and define C_t as the individual's date t rate of consumption per unit time. Also, let ω_i be the proportion of total wealth allocated to risky asset i, $i = 1, ..., n$. Similar to our analysis in Chapter 9 and treating consumption as a net cash outflow from the individual's wealth portfolio, we can write the dynamics of wealth as[2]

$$\begin{aligned} dW &= \left[\sum_{i=1}^{n}\omega_i dS_i/S_i + \left(1-\sum_{i=1}^{n}\omega_i\right)rdt\right]W - Cdt \\ &= \sum_{i=1}^{n}\omega_i(\mu_i - r)W\, dt \;+\; (rW-C)\, dt \;+\; \sum_{i=1}^{n}\omega_i W\sigma_i\, dz_i \end{aligned} \tag{12.3}$$

We can now state the individual's intertemporal consumption and portfolio choice problem:

$$\max_{C_s,\{\,\omega_{i,s}\},\forall s,i} E_t\left[\int_t^T U\left(C_s,s\right)\, ds \;+\; B(W_T,T)\right] \tag{12.4}$$

subject to the constraint (12.3).

[2]Our presentation assumes that there are no other sources of wealth, such as wage income. If the model is extended to include a flow of nontraded wage income received at date t, say, y_t, it could be incorporated into the individual's intertemporal budget constraint in a manner similar to that of consumption but with an opposite sign. In other words, the term $(C_t - y_t)$ would replace C_t in our derivation of the individual's dynamic budget constraint. Duffie, Fleming, Soner, and Zariphopoulou (Duffie, Fleming, Soner, and Zariphopoulou 1997) solve for optimal consumption and portfolio choices when the individual receives stochastic wage income.

The date t utility function, $U(C_t, t)$, is assumed to be strictly increasing and concave in C_t and the bequest function, $B(W_T, T)$, is assumed to be strictly increasing and concave in terminal wealth, W_T. This problem, in which the individual has time-separable utility of consumption, is analogous to the discrete-time problem studied in Chapter 5. The variables W_s and $x(s)$ are the date s state variables while the individual chooses the control variables C_s and $\omega_i(s)$, $i = 1, ..., n$, for each date s over the interval from dates t to T.

Note that some possible constraints have not been imposed. For example, one might wish to impose the constraint $C_t \geq 0$ (nonnegative consumption) and/or $\omega_i \geq 0$ (no short sales). However, for some utility functions, negative consumption is never optimal, so that solutions satisfying $C_t \geq 0$ would result even without the constraint.[3]

Before we attempt to solve this problem, let's digress to consider how stochastic dynamic programming applies to a continuous-time setting.

12.2 Continuous-Time Dynamic Programming

To illustrate the principles of dynamic programming in continuous time, consider a simplified version of the problem specified in conditions (12.3) to (12.4) where there is only one choice variable:

$$\max_{\{c\}} E_t \left[\int_t^T U(c_s, x_s)\, ds \right] \tag{12.5}$$

subject to

$$dx = a(x, c)\, dt + b(x, c)\, dz \tag{12.6}$$

[3] For example, if $\underset{C_t \to 0}{lim} \frac{\partial U(C_t, t)}{\partial C_t} = \infty$, as would be the case if the individual's utility displayed constant relative risk aversion (power utility), then the individual would always avoid nonpositive consumption. However, other utility functions, such as constant absolute-risk-aversion (negative exponential) utility, do not display this property.

where c_t is a *control* variable (such as a consumption and/or vector of portfolio proportions) and x_t is a *state* variable (such as wealth and/or a variable that changes investment opportunities, that is, a variable that affects the μ_i's and/or σ_i's). As in Chapter 5, define the indirect utility function, $J(x_t, t)$, as

$$\begin{aligned} J(x_t,\, t) &= \max_{\{c\}} E_t \left[\int_t^T U(c_s, x_s)\, ds \right] \qquad (12.7) \\ &= \max_{\{c\}} E_t \left[\int_t^{t+\Delta t} U(c_s, x_s)\, ds \;+\; \int_{t+\Delta t}^T U(c_s, x_s)\, ds \right] \end{aligned}$$

Now let us apply Bellman's *Principle of Optimality.* Recall that this concept says that an optimal policy must be such that for a given future realization of the state variable, $x_{t+\Delta t}$, (whose value may be affected by the optimal control policy at date t and earlier), any remaining decisions at date $t+\Delta t$ and later must be optimal with respect to $x_{t+\Delta t}$. In other words, an optimal policy must be time consistent. This allows us to write

$$\begin{aligned} J(x_t,\, t) &= \max_{\{c\}} E_t \left[\int_t^{t+\Delta t} U(c_s, x_s)\, ds \;+\; \max_{\{c\}} E_{t+\Delta t} \left[\int_{t+\Delta t}^T U(c_s, x_s)\, ds \right] \right] \\ &= \max_{\{c\}} E_t \left[\int_t^{t+\Delta t} U(c_s, x_s)\, ds \;+\; J(x_{t+\Delta t},\, t+\Delta t) \right] \qquad (12.8) \end{aligned}$$

Equation (12.8) has the recursive structure of the Bellman equation that we derived earlier in discrete time. However, let us now go a step further by thinking of Δt as a short interval of time and approximate the first integral as $U(c_t, x_t)\,\Delta t$. Also, expand $J(x_{t+\Delta t},\, t+\Delta t)$ around the points x_t and t in a Taylor series to get

$$
\begin{aligned}
J(x_t,\, t) \quad &= \quad \max_{\{c\}} E_t \left[U(c_t, x_t)\,\Delta t \;+\; J(x_t,\, t) \;+\; J_x \Delta x \;+\; J_t \Delta t \right. \\
&\quad \left. +\; \frac{1}{2} J_{xx} (\Delta x)^2 \;+\; J_{xt} (\Delta x)(\Delta t) \;+\; \frac{1}{2} J_{tt} (\Delta t)^2 \;+\; o(\Delta t) \right]
\end{aligned} \tag{12.9}
$$

where $o\,(\Delta t)$ represents higher-order terms, say, $y\,(\Delta t)$, where $\lim_{\Delta t \to 0} \frac{y(\Delta t)}{\Delta t} = 0$. Based on our results from Chapter 8, the state variable's diffusion process (12.6) can be approximated as

$$
\Delta x \;\approx\; a(x,\, c)\Delta t \;+\; b(x,\, c)\Delta z \;+\; o(\Delta t) \tag{12.10}
$$

where $\Delta z = \sqrt{\Delta t}\widetilde{\varepsilon}$ and $\widetilde{\varepsilon} \sim N\,(0,1)$. Substituting (12.10) into (12.9), and subtracting $J(x_t,\, t)$ from both sides, one obtains

$$
0 \;=\; \max_{\{c\}} E_t \left[U(c_t, x_t)\Delta t \;+\; \Delta J \;+\; o(\Delta t) \right] \tag{12.11}
$$

where

$$
\Delta J \;=\; \left[J_t \;+\; J_x a \;+\; \frac{1}{2} J_{xx} b^2 \right] \Delta t \;+\; J_x b \Delta z \tag{12.12}
$$

Equation (12.12) is just a discrete-time version of Itô's lemma. Next, note that in equation (12.11) the term $E_t \left[J_x b \Delta z \right] = 0$ and then divide both sides of (12.11) by Δt. Finally, take the limit as $\Delta t \to 0$ to obtain

$$
0 \;=\; \max_{\{c\}} \left[U(c_t,\, x_t) \;+ J_t \;+\; J_x a \;+\; \frac{1}{2} J_{xx} b^2 \right] \tag{12.13}
$$

which is the stochastic, continuous-time Bellman equation analogous to the discrete time Bellman equation (5.15). Equation (12.13) is sometimes rewritten as

$$
0 \;=\; \max_{\{c\}} \left[U(c_t,\, x_t) \;+\; L[J] \; \right] \tag{12.14}
$$

where $L[\cdot]$ is the *Dynkin operator*. This operator is the "drift" term (expected change per unit time) in $dJ(x,\,t)$ that one obtains by applying Itô's lemma to $J(x,\,t)$. In summary, equation (12.14) gives us a condition that the optimal stochastic control policy, c_t, must satisfy. Let us now return to the complete consumption and portfolio choice problem and apply this solution technique.

12.3 Solving the Continuous-Time Problem

Define the indirect utility-of-wealth function, $J(W,\,x,\,t)$, as

$$J(W, x,\, t) \;=\; \max_{C_s,\{\,\omega_{i,s}\},\forall s,i} E_t\left[\int_t^T U(C_s, s)\, ds \;+\; B(W_T, T)\right] \tag{12.15}$$

and define L as the Dynkin operator with respect to the state variables W and x_i, $i = 1,\,\ldots,\,k$. In other words,

$$\begin{aligned} L\,[J] \;\; = \;\; & \frac{\partial J}{\partial t} \;+\; \left[\sum_{i=1}^{n}\omega_i(\mu_i - r)W + (rW - C)\right]\frac{\partial J}{\partial W} \;+\; \sum_{i=1}^{k} a_i\frac{\partial J}{\partial x_i} \\ & + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}\omega_i\omega_j W^2\frac{\partial^2 J}{\partial W^2} \;+\; \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k} b_{ij}\frac{\partial^2 J}{\partial x_i\,\partial x_j} \\ & + \sum_{j=1}^{k}\sum_{i=1}^{n} W\omega_i\phi_{ij}\frac{\partial^2 J}{\partial W\partial x_j} \end{aligned} \tag{12.16}$$

Thus, using equation (12.14), we have

$$0 \;=\; \max_{C_t,\{\,\omega_{i,t}\}}\left[U(C_t, t) \;+ L[J]\right] \tag{12.17}$$

Given the concavity of U and B, equation (12.17) implies that the optimal choices of C_t and $\omega_{i,t}$ satisfy the conditions we obtain from differentiating $U(C_t,\,t) + L[J]$ and setting the result equal to zero. Hence, the first-order conditions are

$$0 \;=\; \frac{\partial U\left(C^{*},t\right)}{\partial C} \;-\; \frac{\partial J\left(W,x,t\right)}{\partial W} \tag{12.18}$$

$$0 \;=\; W\frac{\partial J}{\partial W}(\mu_i - r) \;+W^2\;\frac{\partial^2 J}{\partial W^2}\sum_{j=1}^{n}\sigma_{ij}\omega_j^{*} \;+\; W\sum_{j=1}^{k}\phi_{ij}\frac{\partial^2 J}{\partial x_j\,\partial W},\quad i=1,\ldots,n \tag{12.19}$$

Equation (12.18) is the envelope condition that we earlier derived in a discrete-time framework as equation (5.19), while equation (12.19) has the discrete-time analog (5.20). Defining the inverse marginal utility function as $G = [\partial U/\partial C]^{-1}$, condition (12.18) can be rewritten as

$$C^{*} = G\left(J_W, t\right) \tag{12.20}$$

where we write J_W as shorthand for $\partial J/\partial W$. Also, the n linear equations in (12.19) can be solved in terms of the optimal portfolio weights. Denote $\Omega \equiv [\sigma_{ij}]$ to be the $n \times n$ instantaneous covariance matrix whose i, j^{th} element is σ_{ij}, and denote the i, j^{th} element of the inverse of Ω to be ν_{ij}; that is, $\Omega^{-1} \equiv [\nu_{ij}]$. Then the solution to (12.19) can be written as

$$\omega_i^{*} = -\frac{J_W}{J_{WW}W}\sum_{j=1}^{n}\nu_{ij}(\mu_j - r) - \sum_{m=1}^{k}\sum_{j=1}^{n}\frac{J_{Wx_m}}{J_{WW}W}\phi_{jm}\nu_{ij},\quad i=1,\ldots,n \tag{12.21}$$

Note that the optimal portfolio weights in (12.21) depend on $-J_W/\left(J_{WW}W\right)$ which is the inverse of relative risk aversion for lifetime utility of wealth.

Given particular functional forms for U and the μ_i's, σ_{ij}'s, and ϕ_{ij}'s, equations (12.20) and (12.21) are functions of the state variables W, x, and derivatives of J, that is, J_W, J_{WW}, and J_{Wx_i}. They can be substituted back into equation (12.17) to obtain a nonlinear partial differential equation (PDE) for J. For some specifications of utility and the processes for asset returns and the state variables, this PDE can be solved to obtain an analytic expression for J

that, in turn, allows for explicit solutions for C^* and the ω_i^* based on (12.20) and (12.21). Examples of such analytical solutions are given in the next two sections. In general, however, one must resort to numerical solutions for J and, therefore, C^* and the ω_i^*.[4]

12.3.1 Constant Investment Opportunities

Let us consider the special case for which asset prices or returns are lognormally distributed, so that continuously compounded rates of return are normally distributed. This occurs when all of the μ_i's (including r) and σ_i's are constants.[5] This means that each asset's expected rate of return and variance of its rate of return do not change; there is a *constant investment opportunity set*. Hence, investment and portfolio choice decisions are independent of the state variables, x, since they do not affect U, B, the μ_i's, or the σ_i's. The only state variable affecting consumption and portfolio choice decisions is wealth, W. This simplifies the above analysis, since now the indirect utility function J depends only on W and t, but not x.

For this constant investment opportunity set case, the optimal portfolio weights in (12.21) simplify to

$$\omega_i^* = -\frac{J_W}{J_{WW}W}\sum_{j=1}^{n}\nu_{ij}(\mu_j - r), \quad i = 1, \ldots, n \tag{12.22}$$

Plugging (12.20) and (12.22) back into the optimality equation (12.17), and

[4] Techniques for solving partial differential equations numerically are covered in Carrier and Pearson (Carrier and Pearson 1976), Judd (Judd 1998), and Rogers and Talay (Rogers and Talay 1997).

[5] Recall that if μ_i and σ_i are constants, then dS_i/S_i follows geometric Brownian motion and $S_i(t) = S_i(0)\,e^{\left(\mu_i - \frac{1}{2}\sigma_i^2\right)t + \sigma_i(z_i(t) - z_i(0))}$ is lognormally distributed over any discrete period since $z_i(t) - z_i(0) \,\tilde{}\, N(0,t)$. Therefore, the *return* on a unit initial investment over this period, $S_i(t)/S_i(0) = e^{\left(\mu_i - \frac{1}{2}\sigma_i^2\right)t + \sigma_i(z_i(t) - z_i(0))}$, is also lognormally distributed. The *continuously-compounded rate of return*, equal to $\ln\left[S_i(t)/S_i(0)\right] = \left(\mu_i - \frac{1}{2}\sigma_i^2\right)t + \sigma_i\left(z_i(t) - z_i(0)\right)$, is normally distributed.

using the fact that $[\nu_{ij}] \equiv \Omega^{-1}$, we have

$$0 = U(G, t) + J_t + J_W(rW - G) - \frac{J_W^2}{2J_{WW}} \sum_{i=1}^{n} \sum_{j=1}^{n} \nu_{ij}(\mu_i - r)(\mu_j - r) \quad (12.23)$$

The nonlinear partial differential equation (12.23) may not have an analytic solution for an arbitrary utility function, U. However, we can still draw some conclusions about the individual's investment behavior by looking at equation (12.22). This expression for the individual's optimal portfolio weights has an interesting implication, but one that might be intuitive given a constant investment opportunity set. Since ν_{ij}, μ_j, and r are constants, the proportion of each risky asset that is optimally held will be proportional to $-J_W/(J_{WW}W)$, which depends only on the total wealth state variable, W. Thus, the proportion of wealth in risky asset i to risky asset k is a constant; that is,

$$\frac{\omega_i^*}{\omega_k^*} = \frac{\sum_{j=1}^{n} \nu_{ij}(\mu_j - r)}{\sum_{j=1}^{n} \nu_{kj}(\mu_j - r)} \quad (12.24)$$

and the proportion of risky asset k to all risky assets is

$$\delta_k = \frac{\omega_k^*}{\sum_{i=1}^{n} \omega_i^*} = \frac{\sum_{j=1}^{n} \nu_{kj}(\mu_j - r)}{\sum_{i=1}^{n} \sum_{j=1}^{n} \nu_{ij}(\mu_j - r)} \quad (12.25)$$

This means that each individual, no matter what her utility function, allocates her portfolio between the risk-free asset, paying return r, and a portfolio of the risky assets that holds the n risky assets in constant proportions, given by (12.25). Hence, two "mutual funds," one holding only the risk-free asset and the other holding a risky-asset portfolio with the weights in (12.25) would satisfy all investors. Only the investor's preferences; current level of wealth, W_t; and the investor's time horizon determine the amounts allocated to the risk-free fund and the risky one.

The implication is that with a constant investment opportunity set, one can think of the investment decision as being just a two-asset decision, where the choice is between the risk-free asset paying rate of return r and a risky asset having expected rate of return μ and variance σ^2 where

$$\mu \equiv \sum_{i=1}^{n} \delta_i \mu_i$$
$$\sigma^2 \equiv \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i \delta_j \sigma_{ij} \tag{12.26}$$

These results are reminiscent of those derived from the single-period mean-variance analysis of Chapter 2. In fact, the relative asset proportions given in (12.24) and (12.25) are exactly the same as those implied by the single-period mean-variance portfolio proportions given in equation (2.42).[6] The instantaneous means and covariances for the continuous-time asset price processes simply replace the previous means and covariances of the single-period multivariate normal asset returns distribution. Again, we can interpret all investors as choosing along an efficient frontier, where the tangency portfolio is given by the weights in (12.25). But what is different in this continuous-time analysis is the assumption regarding the distribution in asset prices. In the discrete-time mean-variance analysis, we needed to assume that asset returns were normally distributed, whereas in the continuous-time context we specified that asset returns were lognormally distributed. This latter assumption is more attractive

[6] Note that the i^{th} element of (2.42) can be written as $w_i^* = \lambda \sum_{j=1}^{n} \nu_{ij} \left(\overline{R}_j - R_f\right)$, which equals (12.22) when $\lambda = -J_W / \left(J_{WW} W\right)$.

since most assets, like bonds and common stocks, have limited liability so that their values cannot become negative. The assumption of a lognormal return distribution embodies this restriction, whereas the assumption of normality does not.

The intuition for why we obtain the single-period Markowitz results in a continuous-time setting with lognormally distributed asset returns is as follows. By allowing continuous rebalancing, an individual's portfolio choice horizon is essentially a very short one; that, is the "period" is instantaneous. Since diffusion processes can be thought of as being instantaneously (locally) normally distributed, our continuous-time environment is as if the individual faces an infinite sequence of similar short portfolio selection periods with normally distributed asset returns.

Let's now look at a special case of the preceding general solution. Specifically, we assume that utility is of the hyperbolic absolute risk aversion (HARA) class.

HARA Utility

Recall from Chapter 1 that HARA utility functions are defined by

$$U(C, t) \; = \; e^{-\rho t} \frac{1-\gamma}{\gamma} \left(\frac{\alpha C}{1-\gamma} + \beta \right)^{\gamma} \tag{12.27}$$

and that this class of utility nests power (constant relative-risk-aversion), exponential (constant absolute risk aversion), and quadratic utility. Robert C. Merton (Merton 1971) derived explicit solutions for this class of utility functions. With HARA utility, optimal consumption given in equation (12.20) becomes

$$C^* \; = \; \frac{1-\gamma}{\alpha} \left[\frac{e^{\rho t} J_W}{\alpha} \right]^{\frac{1}{\gamma-1}} \; - \; \frac{(1-\gamma)\beta}{\alpha} \tag{12.28}$$

and using (12.22) and (12.26), the proportion put in the risky-asset portfolio is

$$\omega^* = -\frac{J_W}{J_{WW}W}\frac{\mu - r}{\sigma^2} \tag{12.29}$$

This solution is incomplete since C^* and ω^* are in terms of J_W and J_{WW}. However, we solve for J in the following manner. Substitute (12.28) and (12.29) into the optimality equation (12.17) or, alternatively, directly simplify equation (12.23) to obtain

$$\begin{aligned} 0 \;=\; & \frac{(1-\gamma)^2}{\gamma} e^{-\rho t}\left[\frac{e^{\rho t}J_W}{\alpha}\right]^{\frac{\gamma}{\gamma-1}} + J_t \\ & + \left(\frac{(1-\gamma)\beta}{\alpha} + rW\right) J_W - \frac{J_W^2}{J_{WW}}\frac{(\mu - r)^2}{2\sigma^2} \end{aligned} \tag{12.30}$$

This is the partial differential equation for J that can be solved subject to a boundary condition for $J(W, T)$. Let us assume a zero bequest function, $B \equiv 0$, so that the appropriate boundary condition is $J(W, T) = 0$. The nonlinear partial differential equation in (12.30) can be simplified by a change in variable $Y = J^{\frac{\gamma}{\gamma-1}}$. This puts it in the form of a Bernoulli-type equation and an analytic solution exists. The expression for the general solution is lengthy and can be found in (Merton 1971). Given this solution for J, one can then calculate J_W to solve for C^* and also calculate J_{WW} to solve for ω^*. It is interesting to note that for this class of HARA utility, C^* is of the form

$$C_t^* = aW_t + b \tag{12.31}$$

and

$$\omega_t^* = g + \frac{h}{W_t} \tag{12.32}$$

where a, b, g, and h are, at most, functions of time. For the special case of constant relative risk aversion where $U(C,t) = e^{-\rho t}C^{\gamma}/\gamma$, the solution is

$$J(W,t) = e^{-\rho t}\left[\frac{1-e^{-a(T-t)}}{a}\right]^{1-\gamma} W^{\gamma}/\gamma \tag{12.33}$$

$$C_t^* = \frac{a}{1-e^{-a(T-t)}}W_t \tag{12.34}$$

and

$$\omega^* = \frac{\mu - r}{(1-\gamma)\sigma^2} \tag{12.35}$$

where $a \equiv \frac{\gamma}{1-\gamma}\left[\frac{\rho}{\gamma} - r - \frac{(\mu-r)^2}{2(1-\gamma)\sigma^2}\right]$. When the individual's planning horizon is infinite, that is, $T \to \infty$, a solution exists only if $a > 0$. In this case, we can see that by taking the limits of equations (12.33) and (12.34) as T becomes infinite, then $J(W,t) = e^{-\rho t}a^{\gamma-1}W^{\gamma}/\gamma$ and consumption is a constant proportion of wealth, $C_t^* = aW_t$.

As mentioned in the previous section, in a continuous-time environment when investment opportunities are constant, we obtain the single-period Markowitz result that an investor will optimally divide her portfolio between the risk-free asset and the tangency portfolio of risky assets given by (12.26). However, this does not imply that an investor with the same form of utility would choose the same portfolio weight in this tangency portfolio for both the continuous-time case and the discrete-time case. Indeed, the optimal portfolio choices can be qualitatively different. In particular, the constant relative-risk-averse individual's optimal portfolio weight (12.35) for the continuous-time case differs from what this individual would choose in the discrete-time Markowitz environment. As covered in an exercise at the end of Chapter 2, an individual with constant relative risk aversion and facing normally distributed risky-asset returns would choose to place his entire portfolio in the risk-free asset; that is, $\omega^* = 0$. The reason is that constant relative-risk-aversion utility is not a defined, real-valued function when end-of-period wealth is zero or negative: marginal utility becomes infinite as end-of-period wealth declines to zero. The implication is that such an investor would avoid assets that have a positive probability of making total end-of-period wealth nonpositive. Because risky assets with normally distributed returns have positive probability of having zero or negative values over a discrete period of time, a constant relative-risk-averse individual who cannot revise her portfolio continuously optimally chooses the corner solution where the entire portfolio consists of the risk-free asset.[7] In contrast, an interior portfolio choice occurs in the continuous-time context where constant investment opportunities imply lognormally distributed returns and a zero bound on the value of risky assets.

A couple of final observations regarding the optimal risky-asset portfolio holding (12.35) are, first, that it is decreasing in the individual's coefficient of relative risk aversion, $(1-\gamma)$. This result is consistent with the received wisdom of financial planners that more risk-averse individuals should choose a smaller portfolio allocation in risky assets. However, the second observation is that this risky-asset allocation is independent of the time horizon, T, which runs counter to the conventional advice that individuals should reduce their allocations in risky assets (stocks) as they approach retirement. An extension of the portfolio choice model that endows individuals with riskless labor income whose present value declines as the individual approaches her retirement is one way of producing the result that the individual should allocate a decreasing proportion of her financial asset portfolio to risky assets.[8] In this case, riskless

[7] This portfolio corner solution result extends to the multiperiod discrete-time environment of Chapter 5. Note that this corner solution does not apply to constant absolute risk aversion where marginal utility continues to be positive and finite even when wealth is nonpositive. This is why portfolio choice models often assume that utility displays constant absolute risk aversion if asset returns are normally distributed.

[8] Zvi Bodie, Robert Merton, and Paul Samuelson (Bodie, Merton, and Samuelson 1992) analyze the effects of labor income on lifetime portfolio choices. John Campbell and Luis Viceira (Campbell and Viceira 2002) provide a broader examination of lifetime portfolio allocation.

human capital (the present value of labor income) is large when the individual is young, and it substitutes for holding the risk-free asset in the individual's financial asset portfolio. As the individual ages, her riskless human capital declines and is replaced by more of the riskless asset in her financial portfolio.

12.3.2 Changing Investment Opportunities

Next, let us generalize the individual's consumption and portfolio choice problem by considering the effects of changing investment opportunities. To keep the analysis fairly simple, assume that there is a single state variable, x. That is, let $k = 1$ so that x is a scalar. We also simplify the notation by writing its process as

$$dx = a\left(x,t\right)dt + b\left(x,t\right)d\zeta \tag{12.36}$$

where $b\, d\zeta \sigma_i \, dz_i = \phi_i \, dt$. This allows us to write the optimal portfolio weights in (12.21) as

$$\omega_i^* = -\frac{J_W}{WJ_{WW}} \sum_{j=1}^{n} \upsilon_{ij}\left(\mu_j - r\right) - \frac{J_{Wx}}{WJ_{WW}} \sum_{j=1}^{n} \upsilon_{ij}\phi_j, \quad i = 1, \ldots, n \tag{12.37}$$

or, written in matrix form,

$$\boldsymbol{\omega}^* = \frac{A}{W}\boldsymbol{\Omega}^{-1}\left(\boldsymbol{\mu} - r\mathbf{e}\right) + \frac{H}{W}\boldsymbol{\Omega}^{-1}\boldsymbol{\phi} \tag{12.38}$$

where $\boldsymbol{\omega}^* = (\omega_1^* ... \omega_n^*)'$ is the $n \times 1$ vector of portfolio weights for the n risky assets; $\boldsymbol{\mu} = (\mu_1 ... \mu_n)'$ is the $n \times 1$ vector of these assets' expected rates of return; $\mathbf{e}$ is an n-dimensional vector of ones, $\boldsymbol{\phi} = (\phi_1, ..., \phi_n)'$, $A = -\frac{J_W}{J_{WW}}$, and $H = -\frac{J_{Wx}}{J_{WW}}$. We will use **bold** type to denote vector or matrix variables, while regular type is used for scalar variables.

Note that A and H will, in general, differ from one individual to another, depending on the form of the particular individual's utility function and level of wealth. Thus, unlike in the constant investment opportunity set case (where $J_{Wx} = H = 0$), ω_i^*/ω_j^* is not the same for all investors, that is; a *two mutual fund theorem* does *not* hold. However, with one state variable, x, a *three fund theorem* does hold. Investors will be satisfied choosing between a fund holding only the risk-free asset, a second fund of risky assets that provides optimal instantaneous diversification, and a third fund composed of a portfolio of the risky assets that has the maximum absolute correlation with the state variable, x. The portfolio weights of the second fund are $\mathbf{\Omega}^{-1}\left(\boldsymbol{\mu} - r\mathbf{e}\right)$ and are the same ones representing the mean-variance efficient tangency portfolio that were derived for the case of constant investment opportunities. The portfolio weights for the third fund are $\mathbf{\Omega}^{-1}\boldsymbol{\phi}$. Note that these weights are of the same form as equation (2.64), which are the hedging demands derived in Chapter 2's cross-hedging example. They are the coefficients from a regression (or projection) of changes in the state variable on the returns of the risky assets A/W and H/W, which depend on the individual's preferences, then determine the relative amounts that the individual invests in the second and third risky portfolios.

To gain more insight regarding the nature of the individual's portfolio holdings, recall the envelope condition $J_W = U_C$, which allows us to write $J_{WW} = U_{CC}\partial C/\partial W$. Therefore, A can be rewritten as

$$A = -\frac{U_C}{U_{CC}\left(\partial C/\partial W\right)} > 0 \tag{12.39}$$

by the concavity of U. Also, since $J_{Wx} = U_{CC}\partial C/\partial x$, we have

$$H = -\frac{\partial C/\partial x}{\partial C/\partial W} \gtreqless 0 \tag{12.40}$$

Now the first vector of terms on the right-hand side of (12.38) represents the usual demand functions for risky assets chosen by a single-period, mean-variance utility maximizer. Since A is proportional to the reciprocal of the individual's absolute risk aversion, we see that the more risk averse the individual, the smaller A is and the smaller in magnitude is the individual's demand for any risky asset.

The second vector of terms on the right-hand side of (12.38) captures the individual's desire to hedge against unfavorable shifts in investment opportunities that would reduce optimal consumption. An unfavorable shift is defined as a change in x such that consumption falls for a given level of current wealth, that is, an increase in x if $\partial C/\partial x < 0$ and a decrease in x if $\partial C/\partial x > 0$. For example, suppose that $\mathbf{\Omega}$ is a diagonal matrix, so that $\upsilon_{ij} = 0$ for $i \neq j$ and $\upsilon_{ii} = 1/\sigma_{ii} > 0$, and also assume that $\phi_i \neq 0$.[9] Then, in this special case, the hedging demand term for risky asset i in (12.38) simplifies to

$$H\upsilon_{ii}\phi_i = -\frac{\partial C/\partial x}{\partial C/\partial W}\upsilon_{ii}\phi_i > 0 \text{ iff } \frac{\partial C}{\partial x}\phi_i < 0 \tag{12.41}$$

Condition (12.41) says that if an increase in x leads to a decrease in optimal consumption ($\partial C/\partial x < 0$) and if x and asset i are positively correlated ($\phi_i > 0$), then there is a positive hedging demand for asset i; that is, $H\upsilon_{ii}\phi_i > 0$ and asset i is held in greater amounts than what would be predicted based on a simple single-period mean-variance analysis. The intuition for this result is that by holding more of asset i, one hedges against a decline in future consumption due to an unfavorable shift in x. If x increases, which would tend to decrease consumption ($\partial C/\partial x < 0$), then asset i would tend to have a high return ($\phi_i > 0$), which by augmenting wealth, W, helps neutralize the fall in consumption ($\partial C/\partial W > 0$). Hence, the individual's optimal portfolio holdings are designed

[9] Alternatively, assume $\mathbf{\Omega}$ is nondiagonal but that $\phi_j = 0$ for $j \neq i$.

to reduce fluctuations in consumption over his planning horizon.

To take a concrete example, suppose that x is a state variable that positively affects the expected rates of return on all assets, including the instantaneously risk-free asset. One simple specification of this is $r = x$ and $\boldsymbol{\mu} = r\mathbf{e}+\mathbf{p} = x\mathbf{e}+\mathbf{p}$ where $\mathbf{p}$ is a vector of risk premia for the risky assets. Thus, an increase in the risk-free rate r indicates an improvement in investment opportunities. Now recall from Chapter 4's equation (4.14) that in a simple certainty model with constant relative-risk-aversion utility, the elasticity of intertemporal substitution is given by $\epsilon = 1/(1-\gamma)$. When $\epsilon < 1$, implying that $\gamma < 0$, it was shown that an increase in the risk-free rate leads to greater current consumption because the income effect is greater than the substitution effect. This result is consistent with equation (12.34) where, for the infinite horizon case of $T \to \infty$, we have $C_t = \frac{\gamma}{1-\gamma}\left[\frac{\rho}{\gamma} - r - \frac{(\mu-r)^2}{2(1-\gamma)\sigma^2}\right] W_t = \frac{\gamma}{1-\gamma}\left[\frac{\rho}{\gamma} - r - \frac{\left(\sum_{i=1}^{n}\delta_i \mathrm{p}_i\right)^2}{2(1-\gamma)\sigma^2}\right] W_t$, so that $\partial C_t/\partial r = -\gamma W_t/(1-\gamma)$.[10] Given empirical evidence that risk aversion is greater than log ($\gamma < 0$), the intuition from these simple models would be that $\partial C_t/\partial r > 0$ and is increasing in risk aversion.

From equation (12.41) we have

$$H\upsilon_{ii}\phi_i = -\frac{\partial C/\partial r}{\partial C/\partial W}\upsilon_{ii}\phi_i > 0 \text{ iff } \frac{\partial C}{\partial r}\phi_i < 0 \tag{12.42}$$

Thus, there is a positive hedging demand for an asset that is negatively correlated with changes in the interest rate, r. An obvious candidate asset would be a bond with a finite time until maturity. For example, if the interest rate followed Vasicek's Ornstein-Uhlenbeck process (Vasicek 1977) given in equation (9.30) of Chapter 9, then any finite-maturity bond whose price process satis-

[10] Technically, it is not valid to infer the derivative $\partial C/\partial r$ from the constant investment opportunities model where we derived optimal consumption assuming r was constant. However, as we shall see from an example later in this chapter, a similar result holds when we solve for optimal consumption using a model where investment opportunities are explicitly changing.

fied equations (9.31) and (9.44) would be perfectly negatively correlated with changes in r. Thus, bonds would be a hedge against adverse changes in investment opportunities since they would experience a positive return when r declines. Moreover, the greater an investor's risk aversion, the greater would be the hedging demand for bonds.

This insight may explain the Asset Allocation Puzzle described by Niko Canner, N. Gregory Mankiw, and David Weil (Canner, Mankiw, and Weil 1997). The puzzle relates to the choice of allocating one's portfolio among three asset classes: stocks, bonds, and cash (where cash refers to a short-maturity money market investment). The conventional wisdom of financial planners is to recommend that an investor hold a lower proportion of her portfolio in stocks and higher proportions in bonds and cash the more risk averse she is. If we consider cash to be the (instantaneous-maturity) risk-free investment paying the return of r, while bonds and stocks are each risky investments, Canner, Mankiw, and Weil point out that this advice is inconsistent with Markowitz's Two-Fund Separation Theorem discussed in Chapter 2. While Markowitz's theory implies that more risk-averse individuals should hold more cash, it also implies that the optimal risky-asset portfolio (tangency portfolio) should be the same for all investors, so that investors' ratio of risky bonds to risky stocks should be identical irrespective of their risk aversions. Therefore, Canner, Mankiw, and Weil conclude that it is puzzling that financial planners recommend a greater bonds-to-stocks mix for more risk-averse investors.

However, based on our previous analysis, specifically equation (12.42), we see that financial planners' advice is consistent with employing bonds as a hedge against changing investment opportunities and that the demand for this hedge increases with an investor's risk aversion. Hence, while the conventional wisdom is inconsistent with static, single-period portfolio rules, it is predicted by

Merton's more sophisticated intertemporal portfolio rules.[11] One caveat with this explanation of the puzzle is that the Merton theory assumes that changing investment opportunities represent real, rather than nominal, variation in asset return distributions. If so, the optimal hedging instrument may be a real (inflation-indexed) bond. In contrast, the asset allocation advice of financial practitioners tends to be in terms of nominal (currency-denominated) bonds. Still, if nominal bond price movements result primarily from changes in real interest rates, rather than expected inflation, then in the absence of indexed bonds, nominal bonds may be the best available hedge against changes in real rates.[12]

The Special Case of Logarithmic Utility

Let us continue to assume that there is a single state variable affecting investment opportunities but now also specify that the individual has logarithmic utility and a logarithmic bequest function, so that in equation (12.15), $U(C_s, s) = e^{-\rho s} \ln (C_s)$ and $B(W_T, T) = e^{-\rho T} \ln (W_T)$. Logarithmic utility is one of the few cases in which analytical solutions for consumption and portfolio choices can be obtained when investment opportunities are changing. To derive the solution to (12.17) for log utility, let us consider a trial solution for the indirect utility function of the form $J(W, x, t) = d(t) U(W_t, t) + F(x, t) = d(t) e^{-\rho t} \ln (W_t) + F(x, t)$. Then optimal consumption in (12.20) would be

$$C_t^* = \frac{W_t}{d(t)} \tag{12.43}$$

[11] Isabelle Bajeux-Besnainou, James Jordan, and Roland Portait (Bajeux-Besnainou, Jordan, and Portait 2001) were among the first to resolve this asset allocation puzzle based on Merton's intertemporal portfolio theory.

[12] In 1997, the year the Canner, Mankiw, and Weil article was published, the United States Treasury began issuing inflation-indexed bonds called Treasury Inflation-Protected Securities (TIPS). Prior to this date, nominal bonds may have been feasible hedges against changing real returns. See research by Michael Brennan and Yihong Xia (Brennan and Xia 2002) and Chapter 3 of the book by John Campbell and Luis Viceira (Campbell and Viceira 2002).

and the first-order conditions for the portfolio weights in (12.37) simplify to

$$\omega_i^* = \sum_{j=1}^{n} \upsilon_{ij} \left(\mu_j - r\right) \tag{12.44}$$

since $J_{Wx} = 0$. Substituting these conditions into the Bellman equation (12.17), it becomes

$$\begin{aligned} 0 &= U\left(C_t^*, t\right) + J_t + J_W\left[rW_t - C_t^*\right] + a\left(x, t\right) J_x \\ &\quad + \frac{1}{2} b\left(x, t\right)^2 J_{xx} - \frac{J_W^2}{2J_{WW}} \sum_{i=1}^{n}\sum_{j=1}^{n} \upsilon_{ij}\left(\mu_j - r\right)\left(\mu_i - r\right) \\ &= e^{-\rho t} \ln\left[\frac{W_t}{d\left(t\right)}\right] + e^{-\rho t}\left[\frac{\partial d\left(t\right)}{\partial t} - \rho d\left(t\right)\right] \ln\left[W_t\right] + F_t + e^{-\rho t} d\left(t\right) r - e^{-\rho t} \\ &\quad + a\left(x, t\right) F_x + \frac{1}{2} b\left(x, t\right)^2 F_{xx} + \frac{d\left(t\right) e^{-\rho t}}{2} \sum_{i=1}^{n}\sum_{j=1}^{n} \upsilon_{ij}\left(\mu_j - r\right)\left(\mu_i - r\right) \end{aligned} \tag{12.45}$$

or

$$\begin{aligned} 0 &= -\ln\left[d\left(t\right)\right] + \left[1 + \frac{\partial d\left(t\right)}{\partial t} - \rho d\left(t\right)\right] \ln\left[W_t\right] + e^{\rho t} F_t + d\left(t\right) r - 1 \\ &\quad + a\left(x, t\right) e^{\rho t} F_x + \frac{1}{2} b\left(x, t\right)^2 e^{\rho t} F_{xx} + \frac{d\left(t\right)}{2} \sum_{i=1}^{n}\sum_{j=1}^{n} \upsilon_{ij}\left(\mu_j - r\right)\left(\mu_i - r\right) \end{aligned} \tag{12.46}$$

Note that a solution to this equation must hold for all values of wealth. Hence, it must be the case that

$$\frac{\partial d\left(t\right)}{\partial t} - \rho d\left(t\right) + 1 = 0 \tag{12.47}$$

subject to the boundary condition $d(T) = 1$. The solution to this first-order ordinary differential equation is

$$d(t) = \frac{1}{\rho}\left[1 - (1-\rho)\, e^{-\rho(T-t)}\right] \tag{12.48}$$

The complete solution to (12.46) is then to solve

$$\begin{aligned} 0 \quad = \quad & -\ln[d(t)] + e^{\rho t} F_t + d(t)\, r - 1 + a(x,t)\, e^{\rho t} F_x \\ & + \frac{1}{2} b(x,t)^2 e^{\rho t} F_{xx} + \frac{d(t)}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \upsilon_{ij} \left(\mu_j - r\right)\left(\mu_i - r\right) \end{aligned} \tag{12.49}$$

subject to the boundary condition $F(x,T) = 0$ and where (12.48) is substituted in for $d(t)$. The solution to (12.49) depends on how r, the μ_i's, and Ω are assumed to depend on the state variable x. However, whatever assumptions are made regarding these variables' relationships to the state variable x, they will influence only the level of indirect utility via the value of $F(x,t)$ and will not change the form of the optimal consumption and portfolio rules. Thus, this verifies that our trial solution is, indeed, a valid form for the solution to the individual's problem. Substituting (12.48) into (12.43), consumption satisfies

$$C_t = \frac{\rho}{1 - (1-\rho)\, e^{-\rho(T-t)}} W_t \tag{12.50}$$

which is the continuous-time counterpart to the log utility investor's optimal consumption that we derived for the discrete-time problem in Chapter 5, equation (5.33). Note also that the log utility investor's optimal portfolio weights given in (12.44) are of the same form as in the case of a constant investment opportunity set, equation (12.35) with $\gamma = 0$. Similar to the discrete-time case, the log utility investor may be described as behaving myopically in that she has no desire to hedge against changes in investment opportunities.[13] However, note that even with log utility, a difference from the constant investment opportunity set case is that since r, the μ_i's, and $\mathbf{\Omega}$ depend, in general, on the constantly changing state variable x_t, the portfolio weights in equation (12.44) vary over time.

Recall that log utility is a very special case and, in general, other utility specifications lead to consumption and portfolio choices that reflect desires to hedge against investment opportunities. An example is given in the next section. After introducing an alternative solution technique, we solve for the consumption and portfolio choices of an individual with general power utility who faces changing investment opportunities.

12.4 The Martingale Approach to Consumption and Portfolio Choice

The preceding sections of this chapter showed how stochastic dynamic programming could be used to find an individual's optimal consumption and portfolio choices. An alternative to this dynamic programming method was developed by John Cox and Chi-Fu Huang (Cox and Huang 1989), Ioannis Karatzas, John Lehoczky, and Steven Shreve (Karatzas, Lehoczky, and Shreve 1987), and Stanley Pliska (Pliska 1986). Their solution technique uses a stochastic discount factor (state price deflator, or pricing kernel) for valuation, and so it is most applicable to an environment characterized by dynamically complete markets.[14]

[13] The portfolio weights for the discrete time case are given by (5.34). As discussed earlier, the log utility investor acts myopically because income and substitution effects from changing investment opportunities exactly cancel for this individual.

[14] Hua He and Neil Pearson (He and Pearson 1991) have extended this martingale approach to an incomplete markets environment. Although in this case there exists an infinity of possible stochastic discount factors, their solution technique chooses what is referred to as a "minimax" martingale measure. This leads to a pricing kernel such that agents do not wish to hedge against the "unhedgeable" uncertainty.

Recall that Chapter 10 demonstrated that when markets are complete, the absence of arbitrage ensures the existence of a unique positive stochastic discount factor. Therefore, let us start by considering the necessary assumptions for market completeness.

12.4.1 Market Completeness Assumptions

As before, let there be n risky assets and a risk-free asset that has an instantaneous return $r(t)$. We modify the previous risky-asset return specification (12.1) to write the return on risky i as

$$dS_i/S_i = \mu_i dt + \mathbf{\Sigma}_i \mathbf{dZ}, \quad i = 1, ..., n \tag{12.51}$$

where $\mathbf{\Sigma}_i = (\sigma_{i1}...\sigma_{in})$ is a $1 \times n$ vector of volatility components and $\mathbf{dZ} = (dz_1...dz_n)'$ is an $n \times 1$ vector of independent Brownian motions.[15] The scalar μ_i, the elements of $\mathbf{\Sigma}_i$, and $r(t)$ may be functions of state variables driven by the Brownian motion elements of $\mathbf{dZ}$. Further, we assume that the n risky assets are nonredundant in the sense that their instantaneous covariance matrix is nonsingular. Specifically, if we let $\mathbf{\Sigma}$ be the $n \times n$ matrix whose i^{th} row equals $\mathbf{\Sigma}_i$, then the instantaneous covariance matrix of the assets' returns, $\mathbf{\Omega} \equiv \mathbf{\Sigma\Sigma}'$, has rank equal to n.

Importantly, we are assuming that any uncertain changes in the means and covariances of the asset return processes in (12.51) are driven only by the vector $\mathbf{dZ}$. This implies that changes in investment opportunities can be perfectly

[15] Note that in (12.51), the independent Brownian motion components of $\mathbf{dZ}$, dz_i, $i = 1, ..., n$ are different from the possibly correlated Brownian motion processes dz_i defined in (12.1). Accordingly, the return on asset i in (12.51) depends on all n of the independent Brownian motion processes, while the return on asset i in (12.1) depends on only one of the correlated Brownian motion processes, namely, the i^{th} one, dz_i. These different ways of writing the risky-asset returns are not important, because an orthogonal transformation of the n correlated Brownian motion processes in (12.1) can allow us to write asset returns as (12.51) where each asset return depends on all n independent processes. The reason for writing asset returns as (12.51) is that individual market prices of risk can be identified with each of the independent risk sources.

hedged by the n assets, and such an assumption makes this market dynamically complete. This differs from the assumptions of (12.1) and (12.2), because we exclude state variables driven by other, arbitrary Brownian motion processes, $d\zeta_i$, that cannot be perfectly hedged by the n assets' returns. Equivalently, if we assume there is a state variable affecting asset returns, say, x_i as represented in (12.2), then its Brownian motion process, $d\zeta_i$, must be a linear function of the Brownian motion components of $\mathbf{dZ}$. Hence, in this section there can be no more than n (not $n+k$) sources of uncertainty affecting the distribution of asset returns.

Given this structure, we showed in Chapter 10 that when arbitrage is not possible, a unique stochastic discount factor exists and follows the process

$$dM/M = -rdt - \mathbf{\Theta}\,(t)'\,\mathbf{dZ} \tag{12.52}$$

where $\mathbf{\Theta} = (\theta_1 ... \theta_n)'$ is an $n \times 1$ vector of market prices of risks associated with each Brownian motion and where $\mathbf{\Theta}$ satisfies

$$\mu_i - r = \mathbf{\Sigma}_i \mathbf{\Theta}, \quad i = 1, ..., n \tag{12.53}$$

Notice that if we take the form of the assets' expected rates of return and volatilities as given, then equation (12.53) is a system of n linear equations that determine the n market prices of risk, $\mathbf{\Theta}$. Alternatively, if $\mathbf{\Theta}$ and the assets' volatilities are taken as given, (12.53) determines the assets' expected rates of return.

12.4.2 The Optimal Consumption Plan

Now consider the individual's original consumption and portfolio choice problem in (12.4) and (12.3). A key to solving this problem is to view the individual's optimally invested wealth as an asset (literally, a portfolio of assets) that pays a continuous dividend equal to the individual's consumption. This implies that the return on wealth, equal to its change in value plus its dividend, can be priced using the stochastic discount factor. The current value of wealth equals the expected discounted value of the dividends (consumption) that it pays over the individual's planning horizon plus discounted terminal wealth.

$$W_t = E_t \left[\int_t^T \frac{M_s}{M_t} C_s ds + \frac{M_T}{M_t} W_T \right] \tag{12.54}$$

Equation (12.54) can be interpreted as an intertemporal budget constraint. This allows the individual's choice of consumption and terminal wealth to be transformed into a static, rather than dynamic, optimization problem. Specifically, the individual's problem can be written as the following Lagrange multiplier problem:[16]

$$\begin{aligned} \max_{C_s \forall s \in [t,T], W_T} E_t & \left[\int_t^T U\left(C_s, s\right) \, ds + B\left(W_T, T\right) \right] \\ & + \lambda \left(M_t W_t - E_t \left[\int_t^T M_s C_s ds + M_T W_T \right] \right) \end{aligned} \tag{12.55}$$

Note that the problem in (12.55) does not explicitly address the portfolio choice decision. This will be determined later by deriving the individual's portfolio trading strategy required to finance his optimal consumption plan.

By treating the integrals in (12.55) as summations over infinite points in time, the first-order conditions for optimal consumption at each date and for terminal wealth are derived as

[16] By specifying the individual's optimal consumption problem as a static constrained optimization, it is straightforward to incorporate additional constraints into the Lagrange multiplier problem. For example, some forms of HARA utility may permit negative consumption. To prevent this, an additional constraint can be added to keep consumption non-negative. For discussion of this issue, see Chapter 6 of Robert Merton's book (Merton 1992).

$$\frac{\partial U\left(C_s,s\right)}{\partial C_s}=\lambda M_s,\quad \forall s\in[t,T] \tag{12.56}$$

$$\frac{\partial B\left(W_T,T\right)}{\partial W_T}=\lambda M_T \tag{12.57}$$

Similar to what we did earlier, define the inverse marginal utility function as $G=[\partial U/\partial C]^{-1}$ and the inverse marginal utility of bequest function as $G_B=[\partial B/\partial W]^{-1}$. This allows us to rewrite these first-order conditions as

$$C_s^*=G\left(\lambda M_s,s\right),\quad \forall s\in[t,T] \tag{12.58}$$

$$W_T^*=G_B\left(\lambda M_T,T\right) \tag{12.59}$$

Except for the yet-to-be-determined Lagrange multiplier λ, equations (12.58) and (12.59) provide solutions to the optimal choices of consumption and terminal wealth. We can now solve for λ based on the condition that the discounted optimal consumption path and terminal wealth must equal the individual's initial endowment of wealth, W_t. Specifically, we substitute (12.58) and (12.59) into (12.54) to obtain

$$W_t=E_t\left[\int_t^T\frac{M_s}{M_t}G\left(\lambda M_s,s\right)ds+\frac{M_T}{M_t}G_B\left(\lambda M_T,T\right)\right] \tag{12.60}$$

Given the initial endowment of wealth, W_t, the distribution of the stochastic discount factor based upon its process in (12.52), and the forms of the utility and bequest functions (which determine G and G_B), the expectation in equation (12.60) can be calculated to determine λ as a function of W_t, M_t, and any date t state variables. Moreover, there is an alternative way to solve for W_t as a function of M_t, λ, and the date t state variables that may sometimes be easier to compute than equation (12.60). As demonstrated in Chapter

10, since wealth represents an asset or contingent claim that pays a dividend equal to consumption, W_t must satisfy a particular Black-Scholes-Merton partial differential equation (PDE) similar to equation (10.7). The equivalence of the stochastic discount factor relationship in (12.60) and this PDE solution was shown to be a result of the assumptions of market completeness and an absence of arbitrage.

To derive the PDE corresponding to (12.60), let us assume for simplicity that there is a single state variable that affects the distribution of asset returns. That is, μ_i, the elements of $\mathbf{\Sigma}_i$, and $r(t)$ may be functions of a single state variable, say, x_t. This state variable follows the process

$$dx = a(x,t)\,dt + \mathbf{B}(x,t)'\,\mathbf{dZ} \tag{12.61}$$

where $\mathbf{B}(x,t) = (B_1...B_n)'$ is an $n \times 1$ vector of volatilities multiplying the Brownian motion components of $\mathbf{dZ}$. Based on (12.60) and the fact that the processes for M_t in (12.52) and x_t in (12.61) are Markov processes, we know that the date t value of optimally invested wealth is a function of M_t and x_t and the individual's time horizon.[17] Hence, by Itô's lemma, the process followed by $W(M_t, x_t, t)$ satisfies

$$\begin{aligned} dW &= W_M dM + W_x dx + \frac{\partial W}{\partial t}dt + \frac{1}{2}W_{MM}(dM)^2 \\ &\quad + W_{Mx}(dM)(dx) + \frac{1}{2}W_{xx}(dx)^2 \\ &= \mu_W dt + \mathbf{\Sigma}_W' \mathbf{dZ} \end{aligned} \tag{12.62}$$

[17] This is because the expectation in (12.60) depends on the distribution of future values of the pricing kernel. From (12.52) and (12.53), the distribution clearly depends on its initial level, M_t, but also on r and $\mathbf{\Theta}$, which can vary with the state variable x.

where

$$\mu_W \equiv -rMW_M + aW_x + \frac{\partial W}{\partial t} + \frac{1}{2}\mathbf{\Theta}'\mathbf{\Theta}M^2W_{MM} - \mathbf{\Theta}'\mathbf{B}MW_{Mx} + \frac{1}{2}\mathbf{B}'\mathbf{B}W_{xx} \tag{12.63}$$

and

$$\mathbf{\Sigma}_W \equiv -W_M M\mathbf{\Theta} + W_x\mathbf{B} \tag{12.64}$$

Following the arguments of Black and Scholes in Chapter 10, the expected return on wealth must earn the instantaneous risk-free rate plus a risk premium, where this risk premium equals the market prices of risk times the sensitivities (volatilities) of wealth to these sources of risk. Specifically,

$$\mu_W + G(\lambda M_t, t) = rW_t + \mathbf{\Sigma}_W'\mathbf{\Theta} \tag{12.65}$$

Wealth's expected return, given by the left-hand side of (12.65), equals the expected change in wealth plus its consumption dividend. Substituting in for μ_W and $\mathbf{\Sigma}_W'$ leads to the PDE

$$\begin{aligned} 0 &= \frac{1}{2}\mathbf{\Theta}'\mathbf{\Theta}M^2W_{MM} - \mathbf{\Theta}'\mathbf{B}MW_{Mx} + \frac{1}{2}\mathbf{B}'\mathbf{B}W_{xx} + \left(\mathbf{\Theta}'\mathbf{\Theta} - r\right)MW_M \\ &\quad + \left(a - \mathbf{B}'\mathbf{\Theta}\right)W_x + \frac{\partial W}{\partial t} + G(\lambda M_t, t) - rW \end{aligned} \tag{12.66}$$

which is solved subject to the boundary condition that terminal wealth is optimal given the bequest motive; that is, $W(M_T, x_T, T) = G_B(\lambda M_T, T)$. Because this PDE is linear, as opposed to the nonlinear PDE for the indirect utility function, $J(W, x, t)$, that results from the dynamic programming approach, it may be relatively easy to solve, either analytically or numerically.

Thus, either equation (12.60) or (12.66) leads to the solution $W(M_t, x_t, t; \lambda) = W_t$ that allows us to determine λ as a function of W_t, M_t, and x_t, and this solution for λ can then be substituted into (12.58) and (12.59). The result is that consumption at any point in time and terminal wealth will depend only on the contemporaneous value of the pricing kernel, that is, $C_s^*(M_s)$ and $W_T^*(M_T)$. Note that when the individual follows this optimal policy, it is time consistent in the sense that should the individual resolve the optimal consumption problem at some future date, say, $s > t$, the computed value of λ will be the same as that derived at date t.

12.4.3 The Portfolio Allocation

Because we have assumed markets are dynamically complete, we know from the results of Chapters 9 and 10 that the individual's optimal process for wealth and its consumption dividend can be replicated by trading in the economy's underlying assets. Thus, our final step is to derive the portfolio allocation policy that finances the individual's consumption and terminal wealth rules. We can do this by comparing the process for wealth in (12.62) to the dynamics of wealth where the portfolio weights in the n risky assets are explicitly represented. Based on the assumed dynamics of asset returns in (12.51), equation (12.3) is

$$\begin{aligned} dW &= \sum_{i=1}^{n} \omega_i(\mu_i - r)W\,dt \;+\; (rW - C_t)\,dt \;+\; W\sum_{i=1}^{n}\omega_i \mathbf{\Sigma}_i \mathbf{dZ} \\ &= \boldsymbol{\omega}'\left(\boldsymbol{\mu} - r\mathbf{e}\right)W\,dt \;+\; (rW - C_t)\,dt + W\boldsymbol{\omega}'\mathbf{\Sigma}\mathbf{dZ} \end{aligned} \tag{12.67}$$

where $\boldsymbol{\omega} = (\omega_1 ... \omega_n)'$ is the $n \times 1$ vector of portfolio weights for the n risky assets and $\boldsymbol{\mu} = (\mu_1 ... \mu_n)'$ is the $n \times 1$ vector of these assets' expected rates of return. Equating the coefficients of the Brownian motion components of the wealth processes in (12.67) and (12.62), we obtain $W\boldsymbol{\omega}'\mathbf{\Sigma} = \mathbf{\Sigma}_W'$. Substituting in (12.64) for $\mathbf{\Sigma}_W$ and rearranging results gives

$$\boldsymbol{\omega} = -\frac{MW_M}{W}\mathbf{\Sigma}'^{-1}\mathbf{\Theta} + \frac{W_x}{W}\mathbf{\Sigma}'^{-1}\mathbf{B} \tag{12.68}$$

Next, recall the no-arbitrage condition (12.53), and note that it can be written in the following matrix form

$$\boldsymbol{\mu} - r\mathbf{e} = \mathbf{\Sigma\Theta} \tag{12.69}$$

Using (12.69) to substitute for $\mathbf{\Theta}$, equation (12.68) becomes

$$\begin{aligned} \boldsymbol{\omega} &= -\frac{MW_M}{W}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}'^{-1}(\boldsymbol{\mu} - r\mathbf{e}) + \frac{W_x}{W}\mathbf{\Sigma}'^{-1}\mathbf{B} \\ &= -\frac{MW_M}{W}\mathbf{\Omega}^{-1}(\boldsymbol{\mu} - r\mathbf{e}) + \frac{W_x}{W}\mathbf{\Sigma}'^{-1}\mathbf{B} \end{aligned} \tag{12.70}$$

These optimal portfolio weights are of the same form as what was derived earlier in (12.38) for the case where the state variable is perfectly correlated with asset returns.[18] A comparison shows that $MW_M = J_W/J_{WW}$ and $W_x = -J_{Wx}/J_{WW}$. Thus, given the solution for $W(M, x, t)$ in (12.60) or (12.66), equation (12.70) represents a derivation of the individual's optimal portfolio choices that is an alternative to the dynamic programming approach. Let us now use this martingale technique to solve a specific consumption-portfolio choice problem.

12.4.4 An Example

An end-of-chapter exercise asks you to use the martingale approach to derive the consumption and portfolio choices for the case of constant investment opportunities and constant relative-risk-aversion utility. As was shown earlier

[18] In this case, $\mathbf{\Omega}^{-1}\boldsymbol{\phi} = \mathbf{\Sigma}'^{-1}\mathbf{B}$.

using the Bellman equation approach, this leads to the consumption and portfolio rules given by equations (12.34) and (12.35). In this section we consider another example analyzed by Jessica Wachter (Wachter 2002) that incorporates changing investment opportunities. A single state variable is assumed to affect the expected rate of return on a risky asset and, to ensure market completeness, this state variable is perfectly correlated with the risky asset's returns. Specifically, let there be a risk-free asset paying a constant rate of return of $r > 0$, and also assume there is a single risky asset so that equation (12.51) can be written simply as

$$dS/S = \mu\left(t\right)dt + \sigma dz \tag{12.71}$$

The risky asset's volatility, σ, is assumed to be a positive constant but the asset's drift is permitted to vary over time. Specifically, let the single market price of risk be $\theta\left(t\right) = \left[\mu\left(t\right) - r\right]/\sigma$. It is assumed to follow the Ornstein-Uhlenbeck process

$$d\theta = a\left(\overline{\theta} - \theta\right)dt - bdz \tag{12.72}$$

where a, $\overline{\theta}$, and b are positive constants. Thus, the market price of risk is perfectly negatively correlated with the risky asset's return.[19] Wachter justifies the assumption of perfect negative correlation as being reasonable based on empirical studies of stock returns. Since $\mu\left(t\right) = r + \theta\left(t\right)\sigma$ and therefore $d\mu = \sigma d\theta$, this model implies that the expected rate of return on the risky asset is mean-reverting, becoming lower (*higher*) after its realized return has been high (*low*).[20]

The individual is assumed to have constant relative-risk-aversion utility and a zero bequest function, so that (12.55) becomes

[19] Robert Merton (Merton 1971) considered a similar problem where the market price of risk was perfectly positively correlated with a risky asset's return.

[20] Straightforward algebra shows that $\mu\left(t\right)$ follows the similar Ornstein-Uhlenbeck process $d\mu = a\left(\overline{\theta}\sigma + r - \mu\right)dt - \sigma bdz$.

$$\max_{C_s \forall s\in[t,T]} E_t\left[\int_t^T e^{-\rho s}\frac{C^\gamma}{\gamma}\,ds\right] + \lambda\left(M_t W_t - E_t\left[\int_t^T M_s C_s ds\right]\right) \tag{12.73}$$

where we have used the fact that it is optimal to set terminal wealth to zero in the absence of a bequest motive. The first-order condition corresponding to (12.58) is then

$$C_s^* = e^{-\frac{\rho s}{1-\gamma}}\left(\lambda M_s\right)^{-\frac{1}{1-\gamma}}, \quad \forall s \in [t,T] \tag{12.74}$$

Therefore, the relationship between current wealth and this optimal consumption policy, equation (12.60), is

$$\begin{aligned} W_t &= E_t\left[\int_t^T \frac{M_s}{M_t} e^{-\frac{\rho s}{1-\gamma}}\left(\lambda M_s\right)^{-\frac{1}{1-\gamma}} ds\right] \\ &= \lambda^{-\frac{1}{1-\gamma}} M_t^{-1} \int_t^T e^{-\frac{\rho s}{1-\gamma}} E_t\left[M_s^{-\frac{\gamma}{1-\gamma}}\right] ds \end{aligned} \tag{12.75}$$

Since $dM/M = -rdt - \theta dz$, the expectation in (12.75) depends only on M_t and the distribution of θ which follows the Ornstein-Uhlenbeck process in (12.72). A solution for W_t can be obtained by computing the expectation in (12.75) directly. Alternatively, one can solve for W_t using the PDE (12.66). For this example, the PDE is

$$\begin{aligned} 0 &= \frac{1}{2}\theta^2 M^2 W_{MM} + \theta b M W_{M\theta} + \frac{1}{2}b^2 W_{\theta\theta} + \left(\theta^2 - r\right) M W_M \\ &\quad + \left[a\left(\overline{\theta} - \theta\right) + b\theta\right] W_\theta + \frac{\partial W}{\partial t} + e^{-\frac{\rho t}{1-\gamma}}\left(\lambda M_t\right)^{-\frac{1}{1-\gamma}} - rW \end{aligned} \tag{12.76}$$

which is solved subject to the boundary condition $W(M_T, \theta_T, T) = 0$ since it is assumed there is no utility from leaving a bequest. Wachter discusses how the equations in (12.75) and (12.76) are similar to ones found in the literature on the term structure of interest rates. When $\gamma < 0$, so that the individual has risk aversion greater than that of a log utility maximizer, the solution is shown to be[21]

$$W_t = (\lambda M_t)^{-\frac{1}{1-\gamma}} e^{-\frac{\rho t}{1-\gamma}} \int_0^{T-t} H(\theta_t, \tau)\, d\tau \tag{12.77}$$

where $H(\theta_t, \tau)$ is the exponential of a quadratic function of θ_t given by

$$H(\theta_t, \tau) \equiv e^{\frac{1}{1-\gamma}\left[A_1(\tau)\frac{\theta_t^2}{2} + A_2(\tau)\theta_t + A_3(\tau)\right]} \tag{12.78}$$

and

$$\begin{aligned} A_1(\tau) &\equiv \frac{2c_1\left(1-e^{-c_3\tau}\right)}{2c_3-(c_2+c_3)\left(1-e^{-c_3\tau}\right)} \\ A_2(\tau) &\equiv \frac{4c_1 a\overline{\theta}\left(1-e^{-c_3\tau/2}\right)^2}{c_3\left[2c_3-(c_2+c_3)\left(1-e^{-c_3\tau}\right)\right]} \\ A_3(\tau) &\equiv \int_0^{\tau}\left[\frac{b^2}{2(1-\gamma)}A_2^2(s) + \frac{b^2}{2}A_1(s) + a\overline{\theta}A_2(s) + \gamma r - \rho\right] ds \end{aligned}$$

with $c_1 \equiv \gamma/(1-\gamma)$, $c_2 \equiv -2(a + c_1 b)$, and $c_3 \equiv \sqrt{c_2^2 - 4c_1 b^2/(1-\gamma)}$. Equation (12.77) can be inverted to solve for the Lagrange multiplier, λ, but since we know from (12.74) that $(\lambda M_t)^{-\frac{1}{1-\gamma}} e^{-\frac{\rho t}{1-\gamma}} = C_t^*$, we can immediately rewrite (12.77) to derive the optimal consumption rule as

$$C_t^* = \frac{W_t}{\int_0^{T-t} H(\theta_t, \tau)\, d\tau} \tag{12.79}$$

The positive function $H(\theta_t, \tau)$ can be given an economic interpretation. Recall that wealth equals the value of consumption from now until $T - t$ periods into the future. Therefore, since $\int_0^{T-t} H(\theta_t, \tau)\, d\tau = W_t/C_t^*$, the function $H(\theta_t, \tau)$ equals the value of consumption τ periods in the future scaled by current con-

[21] This solution also requires $c_2^2 - 4c_1 b^2/(1-\gamma) > 0$.

sumption.

Wachter shows that when $\gamma < 0$ and $\theta_t > 0$, so that the excess return on the risky asset, $\mu(t) - r$, is positive, then $\partial (C_t^* / W_t) / \partial \theta_t > 0$; that is, the individual consumes a greater proportion of wealth the larger the excess rate of return on the risky asset. This is what we would expect given our earlier analysis showing that the "income" effect dominates the "substitution" effect when risk aversion is greater than that of log utility. The higher expected rate of return on the risky asset allows the individual to afford more current consumption, which outweighs the desire to save more in order to take advantage of the higher expected return on wealth.

Let us next solve for this individual's optimal portfolio choice. The risky asset's portfolio weight that finances the optimal consumption plan is given by (12.70) for the case of a single risky asset:

$$\omega = -\frac{MW_M}{W}\frac{\mu(t) - r}{\sigma^2} - \frac{W_\theta}{W}\frac{b}{\sigma} \tag{12.80}$$

Using (12.77), we see that $-MW_M/W = 1/(1-\gamma)$. Moreover, it is straightforward to compute W_θ from (12.77), and by substituting these two derivatives into (12.80) we obtain

$$\begin{aligned} \omega &= \frac{\mu(t) - r}{(1-\gamma)\sigma^2} - \frac{b}{(1-\gamma)\sigma}\frac{\int_0^{T-t} H(\theta_t, \tau)\left[A_1(\tau)\theta_t + A_2(\tau)\right] d\tau}{\int_0^{T-t} H(\theta_t, \tau)\, d\tau} \qquad (12.81) \\ &= \frac{\mu(t) - r}{(1-\gamma)\sigma^2} - \frac{b}{(1-\gamma)\sigma}\int_0^{T-t}\frac{H(\theta_t, \tau)}{\int_0^{T-t} H(\theta_t, \tau')\, d\tau'}\left[A_1(\tau)\theta_t + A_2(\tau)\right] d\tau \end{aligned}$$

The first term is the familiar risky-asset demand whose form is the same as for the case of constant investment opportunities, equation (12.35). The second term on the right-hand side of (12.81) is the demand for hedging against changing investment opportunities. It can be interpreted as a consumption-weighted average of separate demands for hedging against changes in investment opportunities at all horizons from 0 to $T-t$ periods in the future, where the weight at horizon τ is $H(\theta_t,\tau)/\int_0^{T-t} H(\theta_t,\acute{\tau})\,d\acute{\tau}$.

It can be shown that $A_1(\tau)$ and $A_2(\tau)$ are negative when $\gamma < 0$, so that if $\theta_t > 0$, the term $[A_1(\tau)\theta_t + A_2(\tau)]$ is unambiguously negative and, therefore, the hedging demand is positive. Hence, individuals who are more risk averse than log utility place more of their wealth in the risky asset than would be the case if investment opportunities were constant. Because of the negative correlation between risky-asset returns and future investment opportunities, overweighting one's portfolio in the risky asset means that unexpectedly good returns today hedge against returns that are expected to be poorer tomorrow.

12.5 Summary

A continuous-time environment often makes the effects of asset return dynamics on consumption and portfolio decisions more transparent. Interestingly, when asset returns are assumed to be lognormally distributed so that investment opportunities are constant, the individual's optimal portfolio weights are similar in form to those of Chapter 2's single-period mean-variance model that assumed normally distributed asset returns. The fact that the mean-variance optimal portfolio weights could be derived in a multiperiod model with lognormal returns is an attractive result because lognormality is consistent with the limited-liability characteristics of most securities such as bonds and common stocks.

When assets' means and variances are time varying, so that investment opportunities are randomly changing, we found that portfolio allocation rules no longer satisfy the simple mean-variance demands. For cases other than log utility, portfolio choices include additional demand components that reflect a desire to hedge against unfavorable shifts in investment opportunities.

We presented two techniques for finding an individual's optimal consumption and portfolio decisions. The first is a continuous-time analog of the discrete-time dynamic programming approach studied in Chapter 5. This approach leads to a continuous-time Bellman equation, which in turn results in a partial differential equation for the derived utility of wealth. Solving for the derived utility of wealth allows one to then derive the individual's optimal consumption and portfolio choices at each point in time. The second is a martingale solution technique based on the insight that an individual's wealth represents an asset portfolio that pays dividends in the form of a stream of consumption. This permits valuation of the individual's optimal consumption stream using the economy's stochastic discount factor. After deriving the optimal consumption rule, one can then find the portfolio decisions that finance the individual's consumption plan.

This chapter's analysis of an individual's optimal consumption and portfolio decisions provides the foundation for considering the equilibrium returns of assets in a continuous-time economy. This is the topic that we address in the next chapter.

12.6 Exercises

1. Consider the following consumption and portfolio choice problem. An individual must choose between two different assets, a stock and a short (instantaneous) maturity, default-free bond. In addition, the individual faces a stochastic rate of inflation, that is, uncertain changes in the price level (e.g., the Consumer Price Index). The price level (currency price of

the consumption good) follows the process

$$dP_t/P_t = \pi dt + \delta d\zeta$$

The nominal (currency value) of the stock is given by S_t. This nominal stock price satisfies

$$dS_t/S_t = \mu dt + \sigma dz$$

The nominal (currency value) of the bond is given by B_t. It pays an instantaneous nominal rate of return equal to i. Hence, its nominal price satisfies

$$dB_t/B_t = idt$$

Note that $d\zeta$ and dz are standard Wiener processes with $d\zeta dz = \rho dt$. Also assume π, δ, μ, σ, and i are all constants.

a. What processes do the real (consumption good value) rates of return on the stock and the bond satisfy?

b. Let C_t be the individual's date t real rate of consumption and ω be the proportion of real wealth, W_t, that is invested in the stock. Give the process followed by real wealth, W_t.

c. Assume that the individual solves the following problem:

$$\max_{C,\omega} E_0 \int_0^\infty U\left(C_t, t\right) dt$$

subject to the real wealth dynamic budget constraint given in part (b). Assuming $U\left(C_t, t\right)$ is a concave utility function, solve for the individual's optimal choice of ω in terms of the indirect utility-of-wealth function.

d. How does ω vary with ρ? What is the economic intuition for this comparative static result?

2. Consider the individual's intertemporal consumption and portfolio choice problem for the case of a single risky asset and an instantaneously risk-free asset. The individual maximizes expected lifetime utility of the form

$$E_0 \left[\int_0^T e^{-\phi t} u \left(C_t \right) dt \right]$$

The price of the risky asset, S, is assumed to follow the geometric Brownian motion process

$$dS/S = \mu dt + \sigma dz$$

where μ and σ are constants. The instantaneously risk-free asset pays an instantaneous rate of return of r_t. Thus, an investment that takes the form of continually reinvesting at this risk-free rate has a value (price), B_t, that follows the process

$$dB/B = r_t dt$$

where r_t is assumed to change over time, following the Vasicek mean-reverting process (Vasicek 1977)

$$dr_t = a \left[b - r_t \right] dt + s d\zeta$$

where $dz d\zeta = \rho dt$.

a. Write down the intertemporal budget constraint for this problem.

b. What are the two state variables for this consumption-portfolio choice problem? Write down the stochastic, continuous-time Bellman equation for this problem.

c. Take the first-order conditions for the optimal choices of consumption and the demand for the risky asset.

d. Show how the demand for the risky asset can be written as two terms: one term that would be present even if r were constant and another term that exists due to changes in r (investment opportunities).

3. Consider the following resource allocation-portfolio choice problem faced by a university. The university obtains "utility" (e.g., an enhanced reputation for its students, faculty, and alumni) from carrying out research and teaching in two different areas: the "arts" and the "sciences." Let C_a be the number of units of arts activities "consumed" at the university and let C_s be the number of science activities consumed at the university. At date 0, the university is assumed to maximize an expected utility function of the form

$$E_0\left[\int_0^{\infty} e^{-\phi t} u\left(C_a(t), C_s(t)\right) dt\right]$$

where $u\left(C_a, C_s\right)$ is assumed to be increasing and strictly concave with respect to the consumption levels. It is assumed that the cost (or price) of consuming a unit of arts activity is fixed at one. In other words, in what follows we express all values in terms of units of the arts activity, making units of the arts activity the numeraire. Thus, consuming C_a units of the arts activity always costs C_a. The cost (or price) of consuming one unit of science activity at date t is given by $S\left(t\right)$, implying that the university's expenditure on C_s units of science activities costs SC_s. $S\left(t\right)$ is assumed to follow the process

$$dS/S = \alpha_s dt + \sigma_s d\zeta$$

where α_s and σ_s may be functions of S.

The university is assumed to fund its consumption of arts and sciences activities from its endowment. The value of its endowment is denoted W_t. It can be invested in either a risk-free asset or a risky asset. The risk-free asset pays a constant rate of return equal to r. The price of the risky asset is denoted P and is assumed to follow the process

$$dP/P = \mu dt + \sigma dz$$

where μ and σ are constants and $dzd\zeta = \rho dt$. Let ω denote the proportion of the university's endowment invested in the risky asset, and thus $(1 - \omega)$ is the proportion invested in the risk-free asset. The university's problem is then to maximize its expected utility by optimally selecting C_a, C_s, and ω.

a. Write down the university's intertemporal budget constraint, that is, the dynamics for its endowment, W_t.

b. What are the two state variables for this problem? Define a "derived utility of endowment" (wealth) function and write down the stochastic, continuous-time Bellman equation for this problem.

c. Write down the first-order conditions for the optimal choices of C_a, C_s, and ω.

d. Show how the demand for the risky asset can be written as two terms, a standard (single-period) portfolio demand term and a hedging term.

e. For the special case in which utility is given by $u(C_a,C_s) = C_a^{\theta} C_s^{\beta}$, solve for the university's optimal level of arts activity in terms of the level and price of the science activity.

4. Consider an individual's intertemporal consumption, labor, and portfolio choice problem for the case of a risk-free asset and a single risky asset. The individual maximizes expected lifetime utility of the form

$$E_0 \left\{ \int_0^T e^{-\phi t} u(C_t, L_t)\, dt + B(W_T) \right\}$$

where C_t is the individual's consumption at date t and L_t is the amount of labor effort that the individual exerts at date t. $u(C_t, L_t)$ is assumed to be an increasing concave function of C_t but a decreasing concave function of L_t. The risk-free asset pays a constant rate of return equal to r per unit time, and the price of the risky asset, S, satisfies the process

$$dS/S = \mu dt + \sigma dz$$

where μ and σ are constants. For each unit of labor effort exerted at date t, the individual earns an instantaneous flow of labor income of $L_t y_t dt$. The return to effort or wage rate, y_t, is stochastic and follows the process

$$dy = \mu_y(y)\, dt + \sigma_y(y)\, d\zeta$$

where $dz d\zeta = \rho dt$.

a. Letting ω be the proportion of wealth invested in the risky asset, write down the intertemporal budget constraint for this problem.

b. What are the state variables for this problem? Write down the stochastic, continuous-time Bellman equation for this problem.

c. Take the first-order conditions with respect to each of the individual's decision variables.

d. Show how the demand for the risky asset can be written as two terms: one term that would be present even if y were constant and another term that exists due to changes in y.

e. If $u(C_t, L_t) = \gamma \ln [C_t] + \beta \ln [L_t]$, solve for the optimal amount of labor effort in terms of the optimal level of consumption.

5. Consider an individual's intertemporal consumption and portfolio choice problem for the case of two risky assets (with no risk-free asset). The individual maximizes expected lifetime utility of the form

$$E_0 \left\{ \int_0^\infty e^{-\phi t} u(C_t)\, dt \right\}$$

where C_t is the individual's consumption at date t. The individual's portfolio can be invested in a stock whose price, S, follows the process

$$dS/S = \mu dt + \sigma dz$$

and a default-risky bond whose price, B, follows the process

$$dB = rBdt - Bdq$$

where dq is a Poisson counting process defined as

$$dq = \begin{cases} 1 & \text{if a default occurs} \\ 0 & \text{otherwise} \end{cases}$$

The probability of a default occurring over time interval dt is λdt. μ, σ, r, and λ are assumed to be constants. Note that the bond earns a rate of return equal to r when it does not default, but when default occurs, the total amount invested in the bond is lost; that is, the bond price goes to zero, $dB = -B$. We also assume that if default occurs, a new default-risky bond, following the same original bond price process given above, becomes available, so that the individual can always allocate her wealth between the stock and a default-risky bond.

a. Letting ω be the proportion of wealth invested in the stock, write down the intertemporal budget constraint for this problem.

b. Write down the stochastic, continuous-time Bellman equation for this problem. Hint: recall that the Dynkin operator, $L[J]$, reflects the drift terms from applying Itô's lemma to J. In this problem, these terms need to include the expected change in J from jumps in wealth due to bond default.

c. Take the first-order conditions with respect to each of the individual's decision variables.

d. Since this problem reflects constant investment opportunities, it can be shown that when $u(C_t) = c^{\gamma}/\gamma$, $\gamma < 1$, the derived utility-of-wealth function takes the form $J(W,t) = ae^{-\phi t}W^{\gamma}/\gamma$, where a is a positive constant. For this constant relative-risk-aversion case, derive the conditions for optimal C and ω in terms of current wealth and the parameters of the asset price processes. Hint: an explicit formula for ω in terms of all of the other parameters may not be possible because the condition is nonlinear in ω.

e. Maintaining the constant relative-risk-aversion assumption, what is the optimal ω if $\lambda = 0$? Assuming the parameters are such that $0 < \omega < 1$ for this case, how would a small increase in λ affect ω, the proportion of the portfolio held in the stock?

6. Show that a log utility investor's optimal consumption for the continuous time problem, equation (12.50), is comparable to that of the discrete-time problem, equation (5.33).

7. Use the martingale approach to consumption and portfolio choice to solve the following problem. An individual can choose between a risk-free asset paying the interest rate r and a single risky asset whose price satisfies the geometric Brownian motion process

$$\frac{dS}{S} = \mu dt + \sigma dz$$

where r, μ, and σ are constants. This individual's lifetime utility function is time separable, has no bequest function, and displays constant relative risk aversion:

$$E_t \left[\int_t^T e^{-\rho s} \frac{C_s^\gamma}{\gamma} ds \right]$$

a. Assuming an absence of arbitrage, state the form of the market price of risk, θ, in terms of the asset return parameters and write down the process followed by the pricing kernel, dM/M. You need not give any derivations.

b. Write down the individual's consumption choice problem as a static maximization subject to a wealth constraint, where W_t is current wealth and λ is the Lagrange multiplier for the wealth constraint. Derive the first-order conditions for C_s $\forall s \in [t, T]$ and solve for the optimal C_s as a function of λ and M_s.

c. Write down the valuation equation for current wealth, W_t, in terms of λ, M_t, and an integral of expected functions of the future values of the pricing kernel. Given the previous assumptions that the asset price parameters are constants, derive the closed-form solution for this expectation.

d. From the answer in part (c), show that optimal consumption is of the form

$$C_t^* = \frac{a}{1 - e^{-a(T-t)}} W_t$$

where a is a function of r, ρ, γ, and θ.

e. Describe how you next would calculate the optimal portfolio proportion invested in the risky asset, ω, given the results of parts (a) - (d).

Chapter 13

Equilibrium Asset Returns

This chapter considers the equilibrium pricing of assets for a continuous-time economy when individuals have time-separable utility. It derives the Intertemporal Capital Asset Pricing Model (ICAPM) that was developed by Robert Merton (Merton 1973a). One result of this model is to show that the standard single-period CAPM holds for the special case in which investment opportunities are assumed to be constant over time. This is an important modification of the CAPM, not only because the results are extended to a multiperiod environment but because the single-period model's assumption of a normal asset return distribution is replaced with a more attractive assumption of lognormally distributed returns. Since assets such as stocks and bonds have limited liability, the assumption of lognormal returns, which restricts asset values to be nonnegative, is more realistic.

When investment opportunities are changing, the standard "single-beta" CAPM no longer holds. Rather, a multibeta ICAPM is necessary for pricing assets. The additional betas reflect priced sources of risk from additional state variables that affect investment opportunities. However, as was shown by

Douglas Breeden (Breeden 1979), the multibeta ICAPM can be collapsed into a single "consumption" beta model, the so-called Consumption Capital Asset Pricing Model (CCAPM). Thus, consistent with our consumption-based asset pricing results in Chapter 4, the continuous-time, multifactor ICAPM can be interpreted as a consumption-based asset pricing model.

The Merton ICAPM is not a fully general equilibrium analysis because it takes the forms of the assets' return-generating processes as given. However, as this chapter demonstrates, this assumption regarding asset returns can be reconciled with the general equilibrium model of John Cox, Jonathan Ingersoll, and Stephen Ross (CIR) (Cox, Ingersoll, and Ross 1985a). The CIR model is an example of a production economy that specifies the available productive technologies. These technologies are assumed to display constant returns to scale and provide us with a model of asset supplies that is an alternative to the Lucas endowment economy presented in Chapter 6. The CIR framework is useful for determining the equilibrium prices of contingent claims. The final section of this chapter gives an example of how the CIR model can be applied to determine the prices of various maturity bonds that are assumed to be in zero net supply.

13.1 An Intertemporal Capital Asset Pricing Model

Merton's ICAPM is based on the same assumptions made in the previous chapter regarding individuals' consumption and portfolio choices. Individuals can trade in a risk-free asset paying an instantaneous rate of return of $r(t)$ and in n risky assets, where the instantaneous rates of return for the risky assets satisfy

$$dS_i(t) \,/\, S_i(t) \;=\; \mu_i\,(x,t)\; dt \;+\; \sigma_i\,(x,t)\; dz_i \tag{13.1}$$

where $i = 1, ..., n$, and $(\sigma_i \, dz_i)(\sigma_j \, dz_j) = \sigma_{ij} \, dt$. The risk-free return and the means and standard deviations of the risky assets can be functions of time and a $k \times 1$ vector of state variables that follow the processes

$$dx_i = a_i \left(x, t\right) dt + b_i \left(x, t\right) d\zeta_i \tag{13.2}$$

where $i = 1, ..., k$, and $(b_i \, d\zeta_i)(b_j \, d\zeta_j) = b_{ij} \, dt$ and $(\sigma_i \, dz_i)(b_j \, d\zeta_j) = \phi_{ij} \, dt$. Now we wish to consider what must be the equilibrium relationships between the parameters of the asset return processes characterized by equations (13.1) and (13.2). Let us start by analyzing the simplest case first, namely, when investment opportunities are constant through time.

13.1.1 Constant Investment Opportunities

As shown in the previous chapter, when the risk-free rate and the parameters of assets' return processes are constants (r and the μ_i's, σ_i's, and σ_{ij}'s are all constants), the asset price processes in (13.1) are geometric Brownian motions and asset returns are lognormally distributed. In this case, the optimal portfolio choices of all individuals lead them to choose the same portfolio of risky assets. Individuals differ only in how they divide their total wealths between this common risky-asset portfolio and the risk-free asset. For this common risky-asset portfolio, it was shown in Chapter 12's equation (12.25) that the proportion of risky asset k to all risky assets is

$$\delta_k = \frac{\sum_{j=1}^{n} \nu_{kj}(\mu_j - r)}{\sum_{i=1}^{n} \sum_{j=1}^{n} \nu_{ij}(\mu_j - r)} \tag{13.3}$$

and in (12.26) that this portfolio's mean and variance are given by

$$\mu \equiv \sum_{i=1}^{n} \delta_i \mu_i$$
$$\sigma^2 \equiv \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i \delta_j \sigma_{ij}. \tag{13.4}$$

Similar to our derivation of the single-period CAPM, in equilibrium this common risky-asset portfolio must be the market portfolio; that is, $\mu = \mu_m$ and $\sigma^2 = \sigma_m^2$. Moreover, the continuous-time market portfolio is exactly the same as that implied by the single-period CAPM, where the instantaneous means and covariances of the continuous-time asset return processes replace the means and covariances of CAPM's multivariate normal asset return distribution. This implies that the equilibrium asset returns in this continuous-time environment satisfy the same relationship as the single-period CAPM:

$$\mu_i - r = \beta_i \left(\mu_m - r \right), \qquad i = 1, \ . \ . \ ., n \tag{13.5}$$

where $\beta_i \equiv \sigma_{im} / \sigma_m^2$ and σ_{im} is the covariance between the i^{th} asset's rate of return and the market's rate of return. Thus, the constant investment opportunity set assumption replicates the standard, single-period CAPM. Yet, rather than asset returns being normally distributed as in the single-period CAPM, the ICAPM has asset returns being lognormally distributed.

While the standard CAPM results continue to hold for this more realistic intertemporal environment, the assumptions of a constant risk-free rate and unchanging asset return means and variances are untenable. Clearly, interest rates vary over time, as do the volatilities of assets such as common stocks.[1] Moreover, there is substantial evidence that mean returns on assets display

[1] Not only do nominal interest rates vary over time, but there is also evidence that real interest rates do as well (Pennacchi 1991). Also, volatilities of stock returns have been found to follow mean-reverting processes. See, for example, (Bollerslev, Chou, and Kroner 1992) and (Andersen, Bollerslev, Diebold, and Ebens 2001).

predictable time variation.[2] Let us next analyze equilibrium asset pricing for a model that permits such changing investment opportunities.

13.1.2 Stochastic Investment Opportunities

To keep the analysis simple, let us start by assuming that there is a single state variable, x. The system of n equations that a given individual's portfolio weights satisfy is given by the previous chapter's equation (12.19) with $k = 1$. It can be rewritten as

$$0 \;=\; -A(\mu_i - r) \;+\; \sum_{j=1}^{n} \sigma_{ij}\omega_j^* W \;\; - H\phi_i, \quad i = 1, \ldots, n \tag{13.6}$$

where you may recall that $A = -J_W/J_{WW} = -U_C/\left[U_{CC}\left(\partial C/\partial W\right)\right]$ and $H = -J_{Wx}/J_{WW} = -\left(\partial C/\partial x\right)/\left(\partial C/\partial W\right)$. Let's rewrite (13.6) in matrix form, using **bold** type to denote vectors and matrices while using regular type to indicate scalars. Also let the superscript p denote the p^{th} individual's (person's) value of wealth, vector of optimal portfolio weights, and values of A and H. Then (13.6) becomes

$$A^p\left(\boldsymbol{\mu} - r\mathbf{e}\right) = \boldsymbol{\Omega}\boldsymbol{\omega}^{\mathbf{p}} W^p - H^p \boldsymbol{\phi} \tag{13.7}$$

where $\boldsymbol{\mu} = (\mu_1, ..., \mu_n)'$, $\mathbf{e}$ is an n-dimensional vector of ones, $\boldsymbol{\omega}^{\mathbf{p}} = (\omega_1^p, ..., \omega_n^p)'$ and $\boldsymbol{\phi} = (\phi_1, ..., \phi_n)'$. Now if we sum across all individuals and divide both sides by $\sum_p A^p$, we obtain

$$\boldsymbol{\mu} - r\mathbf{e} = a\boldsymbol{\Omega}\boldsymbol{\alpha} - h\boldsymbol{\phi} \tag{13.8}$$

[2] For example, empirical evidence by Narasimhan Jegadeesh and Sheridan Titman (Jegadeesh and Titman 1993) find that abnormal stock returns appear to display positive serial correlation at short horizons up to about a year, a phenomenon described as "momentum." In contrast, there is some evidence (e.g., (Poterba and Summers 1988) and (Fama and French 1988)) that abnormal stock returns are negatively serially correlated over longer-term horizons.

where $a \equiv \sum_p W^p / \sum_p A^p$, $h \equiv \sum_p H^p / \sum_p A^p$, and $\boldsymbol{\alpha} \equiv \sum_p \boldsymbol{\omega^p} W^p / \sum_p W^p$ is the average investment in each asset across investors. These must be the market weights, in equilibrium. Hence, the i^{th} row (i^{th} risky-asset excess return) of equation (13.8) is

$$\mu_i - r = a\sigma_{im} - h\phi_i \tag{13.9}$$

To find the excess return on the market portfolio, we can pre-multiply (13.8) by $\boldsymbol{\alpha}'$ and obtain

$$\mu_m - r = a\sigma_m^2 - h\sigma_{mx} \tag{13.10}$$

where $\sigma_{mx} = \boldsymbol{\alpha}'\boldsymbol{\phi}$ is the covariance between the market portfolio and the state variable, x. Next, define $\boldsymbol{\eta} \equiv \frac{\boldsymbol{\Omega}^{-1}\boldsymbol{\phi}}{\mathbf{e}'\boldsymbol{\Omega}^{-1}\boldsymbol{\phi}}$. By construction, $\boldsymbol{\eta}$ is a vector of portfolio weights for the risky assets, where this portfolio has the maximum absolute correlation with the state variable, x. In this sense, it provides the best possible hedge against changes in the state variable.[3] To find the excess return on this optimal hedge portfolio, we can pre-multiply (13.8) by $\boldsymbol{\eta}'$ and obtain

$$\mu_\eta - r = a\sigma_{\eta m} - h\sigma_{\eta x} \tag{13.11}$$

where $\sigma_{\eta m}$ is the covariance between the optimal hedge portfolio and the market portfolio and $\sigma_{\eta x}$ is the covariance between the optimal hedge portfolio and the state variable, x. Equations (13.10) and (13.11) are two linear equations in the two unknowns, a and h. Solving for a and h and substituting them back into equation (13.9), we obtain:

[3] Note that the numerator of $\boldsymbol{\eta}$, $\boldsymbol{\Omega}^{-1}\boldsymbol{\phi}$, is the $n \times 1$ vector of coefficients from a regression of dx on the n risky-asset returns, dS_i/S_i, $i - 1, ..., n$. Dividing these individual coefficients by their sum, $\mathbf{e}'\boldsymbol{\Omega}^{-1}\boldsymbol{\phi}$, transforms them into portfolio weights.

$$\mu_i - r = \frac{\sigma_{im}\sigma_{\eta x} - \phi_i \sigma_{m\eta}}{\sigma_m^2 \sigma_{\eta x} - \sigma_{mx}\sigma_{m\eta}} \left(\mu_m - r\right) + \frac{\phi_i \sigma_m^2 - \sigma_{im}\sigma_{mx}}{\sigma_m^2 \sigma_{\eta x} - \sigma_{mx}\sigma_{m\eta}} \left(\mu_\eta - r\right) \quad (13.12)$$

While the derivation is somewhat lengthy, it can be shown that (13.12) is equivalent to

$$\begin{aligned} \mu_i - r &= \frac{\sigma_{im}\sigma_\eta^2 - \sigma_{i\eta}\sigma_{m\eta}}{\sigma_m^2 \sigma_\eta^2 - \sigma_{m\eta}^2} \left(\mu_m - r\right) + \frac{\sigma_{i\eta}\sigma_m^2 - \sigma_{im}\sigma_{m\eta}}{\sigma_\eta^2 \sigma_m^2 - \sigma_{m\eta}^2} \left(\mu_\eta - r\right) \\ &\equiv \beta_i^m \left(\mu_m - r\right) + \beta_i^\eta \left(\mu_\eta - r\right) \end{aligned} \quad (13.13)$$

where $\sigma_{i\eta}$ is the covariance between the return on asset i and that of the hedge portfolio. Note that $\sigma_{i\eta} = 0$ if and only if $\phi_i = 0$. For the case in which the state variable, x, is uncorrelated with the market so that $\sigma_{m\eta} = 0$, equation (13.13) simplifies to

$$\mu_i - r = \frac{\sigma_{im}}{\sigma_m^2} \left(\mu_m - r\right) + \frac{\sigma_{i\eta}}{\sigma_\eta^2} \left(\mu_\eta - r\right) \quad (13.14)$$

In this case, the first term on the right-hand side of (13.14) is that found in the standard CAPM. The assumption that x is uncorrelated with the market is not as restrictive as one might first believe, since one could redefine the state variable x as a factor that cannot be explained by current market returns, that is, a factor that is uncorrelated with the market.

An equation such as (13.13) can be derived when more than one state variable exists. In this case, there will be an additional "beta" for each state variable. The intertemporal capital asset pricing relations (ICAPM) given by (13.13) and (13.14) have a form similar to the Arbitrage Pricing Theory of Chapter 3. Indeed, the multifactor ICAPM has been used to justify empirical APT-type factor models. The ICAPM predicts that APT risk factors should be related to changes in investment opportunities. However, it should be noted that, in general, the ICAPM's betas may be time varying and not easy to estimate in a constant-coefficients, multifactor regression model.

13.1.3 An Extension to State-Dependent Utility

It is possible that individuals' utilities may be affected directly by the state of the economy. Here we briefly mention the consequences of allowing the state of nature, x, to influence utility by making it an argument of the utility function, $U(C_t, x_t, t)$. It is straightforward to verify that the form of the individual's continuous-time Bellman equation (12.17), the first order conditions for consumption, C_t, and the portfolio weights, the ω_i's, remain unchanged from those specified in Chapter 12. Hence, our results on the equilibrium returns on assets, equation (13.13), continue to hold. The only change is in the interpretation of H, the individual's hedging demand coefficient. With state-dependent utility, by taking the total derivative of the envelope condition (12.18), one obtains

$$J_{Wx} = U_{CC}\frac{\partial C}{\partial x} + U_{Cx} \tag{13.15}$$

so that

$$H = -\frac{\partial C/\partial x}{\partial C/\partial W} - \frac{U_{Cx}}{U_{CC}\frac{\partial C}{\partial W}} \tag{13.16}$$

It can be shown that, in this case, individuals do not hold portfolios that minimize the variance of consumption. Rather, their portfolio holdings minimize the variance of marginal utility.

13.2 Breeden's Consumption CAPM

Douglas T. Breeden (Breeden 1979) provided a way of simplifying the asset return relationship given in Merton's ICAPM. Breeden's model shows that Chapter 4's single-period consumption-portfolio choice result that an asset's expected rate of return depends upon its covariance with the marginal utility of consumption can be generalized to a multiperiod, continuous-time context.

Breeden considers the same model as Merton and hence, in the case of multiple state variables, derives equation (12.38). Substituting in for A and H, equation (12.38) can be written in matrix form, and for the case of k (multiple) state variables the optimal portfolio weights for the p^{th} investor are given by

$$\boldsymbol{\omega}^{\mathbf{p}} W^p = -\frac{U_C^p}{U_{CC}^p C_W^p}\boldsymbol{\Omega}^{-1}\left(\boldsymbol{\mu} - r\mathbf{e}\right) - \boldsymbol{\Omega}^{-1}\boldsymbol{\Phi}\mathbf{C}_{\mathbf{x}}^{\mathbf{p}}/C_W^p \tag{13.17}$$

where $C_W^p = \partial C^p/\partial W^p$, $\mathbf{C}_{\mathbf{x}}^{\mathbf{p}} = \left(\frac{\partial C^p}{\partial x_1}...\frac{\partial C^p}{\partial x_k}\right)'$, and $\boldsymbol{\Phi}$ is the $n \times k$ matrix of covariances of asset returns with changes in the state variables; that is, its i,j^{th} element is ϕ_{ij}. Pre-multiplying (13.17) by $C_W^p \boldsymbol{\Omega}$ and rearranging terms, we have

$$-\frac{U_C^p}{U_{CC}^p}\left(\boldsymbol{\mu} - r\mathbf{e}\right) = \boldsymbol{\Omega}_{\mathbf{W^p}} C_W^p + \boldsymbol{\Phi}\mathbf{C}_{\mathbf{x}}^{\mathbf{p}} \tag{13.18}$$

where $\boldsymbol{\Omega}_{\mathbf{W^p}}$ is the $n \times 1$ vector of covariances between asset returns with the change in wealth of individual p. Now individual p's optimal consumption, $C^p\left(W^p, \mathbf{x}, t\right)$ is a function of wealth, W^p; the vector of state variables, $\mathbf{x}$; and time, t. Thus, from Itô's lemma, we know that the stochastic terms for dC^p will be

$$C_W^p\left(\omega_1^p W^p \sigma_1 dz_1 + ... + \omega_n^p W^p \sigma_n dz_n\right) + \left(b_1 d\zeta_1 \quad b_2 d\zeta_2 ... b_k d\zeta_k\right)\mathbf{C}_{\mathbf{x}}^{\mathbf{p}} \tag{13.19}$$

Hence, the instantaneous covariances of asset returns with changes in individual p's consumption are given by calculating the instantaneous covariance between each asset (having stochastic term $\sigma_i dz_i$) with the terms given in (13.19). The result, in matrix form, is that the $n \times 1$ vector of covariances between asset returns and changes in the individual's consumption, denoted $\mathbf{\Omega}_{\mathbf{C^p}}$, is

$$\mathbf{\Omega}_{\mathbf{C^p}} = \mathbf{\Omega}_{\mathbf{W^p}} C_W^p + \mathbf{\Phi}\mathbf{C_x^p} \tag{13.20}$$

Note that the right-hand side of (13.20) equals the right-hand side of (13.18), and therefore

$$\mathbf{\Omega}_{\mathbf{C^p}} = -\frac{U_C^p}{U_{CC}^p}\left(\boldsymbol{\mu} - r\mathbf{e}\right) \tag{13.21}$$

Equation (13.21) holds for each individual, p. Next, define C as aggregate consumption per unit time and define $\mathcal{T}$ as an aggregate rate of risk tolerance, where

$$\mathcal{T} \equiv \sum_p -\frac{U_C^p}{U_{CC}^p} \tag{13.22}$$

Then (13.21) can be aggregated over all individuals to obtain

$$\boldsymbol{\mu} - r\mathbf{e} = \mathcal{T}^{-1}\mathbf{\Omega}_{\mathbf{C}} \tag{13.23}$$

where $\mathbf{\Omega}_{\mathbf{C}}$ is the $n \times 1$ vector of covariances between asset returns and changes in aggregate consumption. If we multiply and divide the right-hand side of (13.23) by current aggregate consumption, we obtain

$$\boldsymbol{\mu} - r\mathbf{e} = \left(\mathcal{T}/C\right)^{-1}\mathbf{\Omega}_{\ln \mathbf{C}} \tag{13.24}$$

where $\mathbf{\Omega}_{\ln \mathbf{C}}$ is the $n \times 1$ vector of covariances between asset returns and changes in the logarithm of consumption (percentage rates of change of consumption).

Consider a portfolio, m, with vector of weights $\boldsymbol{\omega}^{\mathbf{m}}$. Pre-multiplying (13.24) by $\boldsymbol{\omega}^{\mathbf{m}\prime}$, we have

$$\mu_m - r = (Ŧ/C)^{-1} \sigma_{m,\ln C} \tag{13.25}$$

where μ_m is the expected return on portfolio m and $\sigma_{m,\ln C}$ is the (scalar) covariance between returns on portfolio m and changes in the log of consumption. Using (13.25) to substitute for $(Ŧ/C)^{-1}$ in (13.24), we have

$$\begin{aligned} \boldsymbol{\mu} - r\mathbf{e} &= (\boldsymbol{\Omega}_{\ln \mathbf{C}}/\sigma_{m,\ln C})(\mu_m - r) \\ &= (\boldsymbol{\beta}_{\mathbf{C}}/\beta_{mC})(\mu_m - r) \end{aligned} \tag{13.26}$$

where $\boldsymbol{\beta}_{\mathbf{C}}$ and β_{mC} are the "consumption betas" of asset returns and of portfolio m's return. The consumption beta for any asset is defined as

$$\beta_{iC} = cov\,(dS_i/S_i, d\ln C)\,/var\,(d\ln C) \tag{13.27}$$

Portfolio m may be any portfolio of assets, not necessarily the market portfolio. Equation (13.26) says that the ratio of expected excess returns on any two assets or portfolios of assets is equal to the ratio of their betas measured relative to aggregate consumption. Hence, the risk of a security's return can be summarized by a single consumption beta. Aggregate optimal consumption, $C(W, \mathbf{x}, t)$, encompasses the effects of levels of wealth and the state variables and in this way is a sufficient statistic for the value of asset returns in different states of the world.

Breeden's consumption CAPM (CCAPM) is a considerable simplification relative to Merton's multibeta ICAPM. Furthermore, while the multiple state variables in Merton's model may not be directly identified or observed, and hence the multiple state variable "betas" may not be computed, Breeden's consumption beta can be computed given that we have data on aggregate consumption. However, as discussed earlier, the results of empirical tests using aggregate consumption data are unimpressive.[4] As in all of our earlier asset pricing models based on individuals' optimal consumption and portfolio choices, the CCAPM and ICAPM rely on the assumption of time-separable utility. When we depart from this restriction on utility, as we do in the next chapter, consumption-based models are able to better describe empirical distributions of asset prices.

The ICAPM and CCAPM are not general equilibrium models in a strict sense. While they model individuals' "tastes" by specifying the form of their utilities, they do not link the asset return processes in (13.1) and (13.2) to the economy's "technologies." A fully general equilibrium model would not start by specifying these assets' return processes but, rather, by specifying the economy's physical production possibilities. In other words, it would specify the economy's productive opportunities that determine the supplies of assets in the economy. By matching individuals' asset demands with the asset supplies, the returns on assets would then be determined endogenously. The Lucas endowment economy model in Chapter 6 was an example of this, and we now turn to another general equilibrium model, namely, Cox, Ingersoll, and Ross's production economy model.

[4] An exception is research by Martin Lettau and Sydney Ludvigson (Lettau and Ludvigson 2001), who find that the CCAPM is successful in explaining stock returns when the model's parameters are permitted to vary over time with the log consumption-wealth ratio.

13.3 A Cox, Ingersoll, and Ross Production Economy

In two companion articles (Cox, Ingersoll, and Ross 1985a);(Cox, Ingersoll, and Ross 1985b), John Cox, Jonathan Ingersoll, and Stephen Ross (CIR) developed a continuous-time model of a production economy that is a general equilibrium framework for many of the asset pricing results of this chapter. Their model starts from basic assumptions regarding individuals' preferences and the economy's production possibilities. Individuals are assumed to have identical preferences and initial wealth as well as to maximize standard, time-separable utility similar to the lifetime utility previously specified in this and the previous chapter, namely, in (12.4).[5] The unique feature of the CIR model is the economy's technologies.

Recall that in the general equilibrium endowment economy model of Robert Lucas (Lucas 1978), technologies are assumed to produce perishable output (dividends) that could not be reinvested, only consumed. In this sense, these Lucas technologies are inelastically supplied. Individuals cannot save output and physically reinvest it to increase the productive capacity of the economy. Rather, in the Lucas economy, prices of the technologies adjust endogenously to make investors' changing demands equal to the technologies' fixed supplies. Given the technologies' distribution of future output (dividends), these prices determine the technologies' equilibrium rates of return.

In contrast, the CIR production economy makes the opposite assumption regarding the supply of technologies. Technologies are in perfectly elastic supply. Individuals can save some of the economy's output and reinvest it, thereby changing the productive capacity of the economy. Assets' rates of return are

[5] When individuals are assumed to have the same utility and initial wealth, we can think of there being a "representative" individual.

pinned down by the economy's technologies' rates of return, and the amounts invested in these technologies become endogenous.

Specifically, CIR assumes that there is a single good that can be either consumed or invested. This "capital-consumption" good can be invested in any of n different risky technologies that produce an instantaneous change in the amount of the consumption good. If an amount η_i is physically invested in technology i, then the proportional change in the amount of this good that is produced is given by

$$\frac{d\eta_i(t)}{\eta_i(t)} = \mu_i(x,t)\, dt + \sigma_i(x,t)\, dz_i, \quad i = 1, ..., n \tag{13.28}$$

where $(\sigma_i\, dz_i)(\sigma_j\, dz_j) = \sigma_{ij}\, dt$. μ_i is the instantaneous expected rate of change in the amount of the invested good and σ_i is the instantaneous standard deviation of this rate of change. Note that because μ_i and σ_i are independent of η_i, the change in the quantity of the good is linear in the amount invested. Hence, each technology is characterized by "constant returns to scale." μ_i and σ_i can vary with time and with a $k \times 1$ vector of state variables, $x(t)$. Thus, the economy's technologies for transforming consumption into more consumption can reflect changing (physical) investment opportunities. The i^{th} state variable is assumed to follow the process

$$dx_i = a_i(x,t)\, dt + b_i(x,t)\, d\zeta_i \tag{13.29}$$

where $i = 1, ..., k$, and $(b_i\, d\zeta_i)(b_j\, d\zeta_j) = b_{ij}\, dt$ and $(\sigma_i\, dz_i)(b_j\, d\zeta_j) = \phi_{ij}\, dt$.

Note that equations (13.28) and (13.29) are nearly identical to our earlier modeling of financial asset returns, equations (13.1) and (13.2). Whereas $dS_i(t)/S_i(t)$ in (13.1) represented a security's proportional return, $d\eta_i(t)/\eta_i(t)$ in (13.28) represents a physical investment's proportional return. However, if each technology is interpreted as being owned by an individual firm, and each of these firms is financed entirely by shareholders' equity, then the rate of return on shareholders' equity of firm i, $dS_i(t)/S_i(t)$, equals the proportional change in the value of the firm's physical assets (capital), $d\eta_i(t)/\eta_i(t)$. Here, $dS_i(t)/S_i(t) = d\eta_i(t)/\eta_i(t)$ equals the instantaneous dividend yield where dividends come in the form of a physical capital-consumption good.

Like the Lucas endowment ecomony, we can think of the CIR production economy as arising from a set of production processes that pay physical dividends. The difference is that the Lucas economy's dividend is in the form of a consumption-only good, whereas the CIR economy's dividend is a capital-consumption good that can be physically reinvested to expand the capacities of the productive output processes. The CIR representative individuals must decide how much of their wealth (the capital-consumption good) to consume versus save and, of the amount saved, how to allocate it between the n different technologies (or firms).

Because equations (13.28) and (13.29) model an economy's production possibilities as constant returns-to-scale technologies, the distributions of assets' rates of return available to investors are exogenous. In one sense, this situation is not different from our earlier modeling of an investor's optimal consumption and portfolio choices. However, CIR's specification allows one to solve for the equilibrium prices of securities other than those represented by the n risky technologies. This is done by imagining there to be other securities that have zero net supplies. For example, there may be no technology that produces an instantaneously risk-free return; that is, $\sigma_i \neq 0 \ \forall \ i$. However, one can solve for the equilibrium riskless borrowing or lending rate, call it $r(t)$, for which the representative individuals would be just indifferent between borrowing or lending. In other words, r would be the riskless rate such that individuals choose to invest zero amounts of the consumption good at this rate. Since all individuals are identical, this amounts to the riskless investment having a zero supply in the economy, so that r is really a "shadow" riskless rate. Yet, this rate would be consistent, in equilibrium, with the specification of the economy's other technologies.

Let us solve for this equilibrium riskless rate in the CIR economy. The individual's consumption and portfolio choice problem is similar to that in Chapter 12, (12.4), except that the individual's savings are now allocated, either directly or indirectly through firms, to the n technologies. An equilibrium is defined as a set of interest rate, consumption, and portfolio weight processes $\{r, C^*, \omega_1^*, ..., \omega_n^*\}$ such that the representative individual's first order conditions hold and markets clear: $\sum_{i=1}^n \omega_i = 1$ and $\omega_i \geq 0$. Note that because $\sum_{i=1}^n \omega_i = 1$, this definition of equilibrium implies that riskless borrowing and lending at the equilibrium rate r has zero net supply. Further, since the capital-consumption good is being physically invested in the technological processes, the constraint against short-selling, $\omega_i \geq 0$, applies.

To solve for the representative individual's optimal consumption and portfolio weights, note that since in equilibrium the individual does not borrow or lend, the individual's situation is exactly as if a riskless asset did not exist. Hence, the individual's consumption and portfolio choice problem is the same one as in the previous chapter but where the process for wealth excludes a risk-free asset. Specifically, the individual solves

$$\max_{C_s, \{\omega_{i,s}\}, \forall s,i} E_t \left[\int_t^T U(C_s, s)\, ds \; + \; B(W_T, T) \right] \tag{13.30}$$

subject to

$$dW \;=\; \sum_{i=1}^n \omega_i W \mu_i \, dt \;-\; C_t \, dt \;+\; \sum_{i=1}^n \omega_i W \sigma_i \, dz_i \tag{13.31}$$

and also subject to the condition $\sum_{i=1}^{n} \omega_i = 1$ and the constraint that $\omega_i \geq 0$. The individual's first-order condition for consumption is the usual one:

$$0 = \frac{\partial U\left(C^*, t\right)}{\partial C} - \frac{\partial J\left(W, x, t\right)}{\partial W} \tag{13.32}$$

but the first-order conditions with respect to the portfolio weights are modified slightly. If we let λ be the Lagrange multiplier associated with the equality $\sum_{i=1}^{n} \omega_i = 1$, then the appropriate first-order conditions for the portfolio weights are

$$\begin{aligned} \Psi_i &\equiv \frac{\partial J}{\partial W}\mu_i W + \frac{\partial^2 J}{\partial W^2}\sum_{j=1}^{n}\sigma_{ij}\omega_j^* W^2 + \sum_{j=1}^{k}\frac{\partial^2 J}{\partial x_j\,\partial W}\phi_{ij}W - \lambda \leq 0 \\ 0 &= \Psi_i \omega_i^* \qquad\qquad i = 1, \ldots, n \end{aligned} \tag{13.33}$$

The Kuhn-Tucker conditions in (13.33) imply that if $\Psi_i < 0$, then $\omega_i^* = 0$, so that in this case the i^{th} technology would not be employed. Assuming that the parameters in (13.28) and (13.29) are such that all technologies are employed, that is, $\Psi_i = 0\ \forall i$, then the solution to the system of equations in (13.33) is

$$\omega_i^* = -\frac{J_W}{J_{WW}W}\sum_{j=1}^{n}\nu_{ij}\mu_j - \sum_{m=1}^{k}\sum_{j=1}^{n}\frac{J_{Wx_m}}{J_{WW}W}\nu_{ij}\phi_{jm} + \frac{\lambda}{J_{WW}W^2}\sum_{j=1}^{n}\nu_{ij} \tag{13.34}$$

for $i = 1, ..., n$. Using our previously defined matrix notation, (13.34) can be rewritten as

$$\boldsymbol{\omega}^* = \frac{A}{W}\boldsymbol{\Omega}^{-1}\boldsymbol{\mu} - \frac{A\lambda}{J_W W^2}\boldsymbol{\Omega}^{-1}\mathbf{e} + \sum_{j=1}^{k}\frac{H_j}{W}\boldsymbol{\Omega}^{-1}\boldsymbol{\phi}_j \tag{13.35}$$

where $A = -J_W/J_{WW}$, $H_j = -J_{Wx_j}/J_{WW}$, and $\boldsymbol{\phi}_j = (\phi_{1j}, ..., \phi_{nj})'$. These portfolio weights can be interpreted as a linear combination of $k + 2$ portfolios. The first two portfolios are mean-variance efficient portfolios in a single-period, Markowitz portfolio selection model: $\mathbf{\Omega}^{-1}\boldsymbol{\mu}$ is the portfolio on the efficient frontier that is tangent to a line drawn from the origin (a zero interest rate) while $\mathbf{\Omega}^{-1}\mathbf{e}$ is the global minimum variance portfolio.[6] The last k portfolios, $\mathbf{\Omega}^{-1}\boldsymbol{\phi}_j$, $j = 1, ..., k$, are held to hedge against changes in the technological risks (investment opportunities). The proportions of these $k+2$ portfolios chosen depend on the individual's utility. An exact solution is found in the usual manner of substituting (13.35) and (12.18) into the Bellman equation. For specific functional forms, a value for the indirect utility function, $J(W, x, t)$ can be derived. This, along with the restriction $\sum_{i=1}^{n} \omega_i = 1$, allows the specific optimal consumption and portfolio weights to be determined.

Since in the CIR economy the riskless asset is in zero net supply, we know that the portfolio weights in (13.35) must be those chosen by the representative individual even if offered the opportunity to borrow or lend at rate r. Recall from the previous chapter's equation (12.21) that these conditions, rewritten in matrix notation, are

$$\boldsymbol{\omega}^* = \frac{A}{W}\mathbf{\Omega}^{-1}(\boldsymbol{\mu} - r\mathbf{e}) + \sum_{j=1}^{k} \frac{H_j}{W}\mathbf{\Omega}^{-1}\boldsymbol{\phi}_j, \quad i = 1, \ldots, n \tag{13.36}$$

Since the individual takes prices and rates as given, the portfolio choices given by the first-order conditions in (13.36) namely, the case when a riskless asset exists therefore must be the same as (13.35). By inspection, the weights in (13.35) and (13.36) are identical when $r = \lambda/(J_W W)$. Hence, substituting for λ in terms of the optimal portfolio weights, we can write the equilibrium interest

[6] Recall that a linear combination of any two portfolios on the mean-variance frontier can create any other portfolio on the frontier.

as[7]

$$
\begin{aligned}
r &= \frac{\lambda}{W J_W} \\
&= \boldsymbol{\omega}^{*\prime}\boldsymbol{\mu} - \frac{W}{A}\boldsymbol{\omega}^{*\prime}\boldsymbol{\Omega}\boldsymbol{\omega}^{*} + \sum_{j=1}^{k}\frac{H_j}{A}\boldsymbol{\omega}^{*\prime}\boldsymbol{\phi}_j
\end{aligned}
\tag{13.37}
$$

Note that equation (13.37) is the same as the previously derived relationship (13.10) except that (13.37) is extended to k state variables. Hence, Merton's ICAPM, as well as Breeden's CCAPM, hold for the CIR economy.

The CIR model also can be used to find the equilibrium shadow prices of other securities that are assumed to have zero net supplies. Such "contingent claims" could include securities such as longer maturity bonds or options and futures. For example, suppose a zero-net-supply contingent claim has a payoff whose value could depend on wealth, time, and the state variables, $P\left(W, t, \{x_i\}\right)$.[8] Itô's lemma implies that its price will follow a process of the form

$$
dP = uPdt + P_W W \sum_{i=1}^{n} \omega_i^* \sigma_i dz_i + \sum_{i=1}^{k} P_{x_i} b_i d\zeta_i \tag{13.38}
$$

where

$$
\begin{aligned}
uP &= P_W\left(W\boldsymbol{\omega}^{*\prime}\boldsymbol{\mu} - C\right) + \sum_{i=1}^{k} P_{x_i} a_i + P_t + \frac{P_{WW}W^2}{2}\boldsymbol{\omega}^{*\prime}\boldsymbol{\Omega}\boldsymbol{\omega}^{*} \\
&\quad + \sum_{i=1}^{k} P_{Wx_i} W \boldsymbol{\omega}^{*\prime}\boldsymbol{\phi}_i + \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k} P_{x_i x_j} b_{ij}
\end{aligned}
\tag{13.39}
$$

Using the Merton ICAPM result (13.9) extended to k state variables, the ex-

[7] To derive the second line in (13.37), it is easiest to write in matrix form the first-order condtions in (13.33) and assume these conditions all hold as equalities. Then solve for λ by pre-multiplying by $\boldsymbol{\omega}^{*\prime}$ and noting that $\boldsymbol{\omega}^{*\prime}\mathbf{e} = 1$.

[8] A contingent claim whose payoff depends on the returns or prices of the technologies can be found by the Black-Scholes methodolgy described in Chapter 9.

pected rate of return on the contingent claim must also satisfy[9]

$$u = r + \frac{W}{A} Cov\left(dP/P, dW/W\right) - \sum_{i=1}^{k} \frac{H_i}{A} Cov\left(dP/P, dx_i\right) \tag{13.40}$$

or

$$\begin{aligned} uP &= rP + \frac{1}{A} Cov\left(dP, dW\right) - \sum_{i=1}^{k} \frac{H_i}{A} Cov\left(dP, dx_i\right) \\ &= rP + \frac{1}{A}\left(P_W W^2 \boldsymbol{\omega}^{*\prime} \boldsymbol{\Omega} \boldsymbol{\omega}^* + \sum_{i=1}^{k} P_{x_i} W \boldsymbol{\omega}^{*\prime} \boldsymbol{\phi}_i\right) \\ &\quad - \sum_{i=1}^{k} \frac{H_i}{A}\left(P_W W \boldsymbol{\omega}^{*\prime} \boldsymbol{\phi}_i + \sum_{j=1}^{k} P_{x_j} b_{ij}\right) \end{aligned} \tag{13.41}$$

where in (13.40) we make use of the fact that the market portfolio equals the optimally invested wealth of the representative individual. Equating (13.39) and (13.41) and recalling the value of the equilibrium risk-free rate in (13.37), we obtain a partial differential equation for the contingent claim's value:[10]

$$\begin{aligned} 0 \;=\; & \frac{P_{WW} W^2}{2} \boldsymbol{\omega}^{*\prime} \boldsymbol{\Omega} \boldsymbol{\omega}^* + \sum_{i=1}^{k} P_{W x_i} W \boldsymbol{\omega}^{*\prime} \boldsymbol{\phi}_i + \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} P_{x_i x_j} b_{ij} + P_t + \\ & P_W\left(rW - C\right) + \sum_{i=1}^{k} P_{x_i}\left[a_i - \frac{W}{A} \boldsymbol{\omega}^{*\prime} \boldsymbol{\phi}_i + \sum_{j=1}^{k} \frac{H_j b_{ij}}{A}\right] - rP \end{aligned} \tag{13.42}$$

The next section illustrates how (13.37) and (13.42) can be used to find the risk free rate and particular contingent claims for a specific case of a CIR

[9] Condition (13.9) can be derived for the case of a contingent claim by using the fact that the contingent claim's weight in the market portfolio is zero.

[10] It is straightforward to derive the valuation equation for a contingent claim that pays a continuous dividend at rate $\delta\left(W, x, t\right) dt$. In this case, the additional term $\delta\left(W, x, t\right)$ appears on the right-hand side of equation (13.42).

economy.

13.3.1 An Example Using Log Utility

The example in this section is based on (Cox, Ingersoll, and Ross 1985b). It assumes that the representative individual's utility and bequest functions are logarithmic and of the form $U(C_s, s) = e^{-\rho s} \ln (C_s)$ and $B(W_T, T) = e^{-\rho T} \ln (W_T)$. For this specification, we showed in the previous chapter that the indirect utility function was separable and equaled $J(W, x, t) = d(t) e^{-\rho t} \ln (W_t) + F(x, t)$ where $d(t) = \frac{1}{\rho}\left[1 - (1-\rho) e^{-\rho(T-t)}\right]$, so that optimal consumption satisfies equation (12.50) and the optimal portfolio proportions equal (12.44). Since $J_{Wx_i} = 0$, $H_i = 0$, and $A = W$, the portfolio proportions in (13.35) simplify to

$$\boldsymbol{\omega}^* = \boldsymbol{\Omega}^{-1} (\boldsymbol{\mu} - r\mathbf{e}) \tag{13.43}$$

where we have used the result that $r = \lambda / (J_W W)$. Using the market clearing condition $\mathbf{e}'\boldsymbol{\omega}^* = 1$, we can solve for the equilibrium risk-free rate:

$$r = \frac{\mathbf{e}'\boldsymbol{\Omega}^{-1}\boldsymbol{\mu} - 1}{\mathbf{e}'\boldsymbol{\Omega}^{-1}\mathbf{e}} \tag{13.44}$$

Substituting (13.44) into (13.43), we see that the optimal portfolio weights are

$$\boldsymbol{\omega}^* = \boldsymbol{\Omega}^{-1}\left[\boldsymbol{\mu} - \left(\frac{\mathbf{e}'\boldsymbol{\Omega}^{-1}\boldsymbol{\mu} - 1}{\mathbf{e}'\boldsymbol{\Omega}^{-1}\mathbf{e}}\right)\mathbf{e}\right] \tag{13.45}$$

Let us next assume that a single state variable, $x(t)$, affects all production processes in the following manner:

$$d\eta_i / \eta_i = \widehat{\mu}_i x \, dt + \widehat{\sigma}_i \sqrt{x} dz_i, \quad i = 1, ..., n \tag{13.46}$$

where $\widehat{\mu}_i$ and $\widehat{\sigma}_i$ are assumed to be constants and the state variable follows the *square root process*[11]

$$dx = (a_0 + a_1 x)\, dt + b_0 \sqrt{x} d\zeta \tag{13.47}$$

where $dz_i d\zeta = \rho_i dt$. Note that this specification implies that the means and variances of the technologies' rates of return are proportional to the state variable. If $a_0 > 0$ and $a_1 < 0$, x is a nonnegative, mean-reverting random variable. A rise in x raises all technologies' expected rates of return but also increases their variances.

We can write the technologies' $n \times 1$ vector of expected rates of return as $\boldsymbol{\mu} = \widehat{\boldsymbol{\mu}} x$ and their $n \times n$ matrix of rate of return covariances as $\boldsymbol{\Omega} = \widehat{\boldsymbol{\Omega}} x$. Using these distributional assumptions in (13.44), we find that the equilibrium interest rate is proportional to the state variable:

$$r = \frac{\mathbf{e}' \widehat{\boldsymbol{\Omega}}^{-1} \widehat{\boldsymbol{\mu}} - 1}{\mathbf{e}' \widehat{\boldsymbol{\Omega}}^{-1} \mathbf{e}} x = \alpha x \tag{13.48}$$

where $\alpha \equiv \left(\mathbf{e}' \widehat{\boldsymbol{\Omega}}^{-1} \widehat{\boldsymbol{\mu}} - 1 \right) / \mathbf{e}' \widehat{\boldsymbol{\Omega}}^{-1} \mathbf{e}$ is a constant. This implies that the risk-free rate follows a square root process of the form

$$dr = \alpha dx = \kappa \left(\overline{r} - r \right) dt + \sigma \sqrt{r} d\zeta \tag{13.49}$$

where $\kappa \equiv -a_1 > 0$, $\overline{r} \equiv -\alpha a_0 / a_1 > 0$, and $\sigma \equiv b_0 \sqrt{\alpha}$. CIR (Cox, Ingersoll, and Ross 1985b) state that when the parameters satisfy $2\kappa \overline{r} \geq \sigma^2$, then if $r(t)$ is currently positive, it will remain positive at all future dates $T \geq t$. This is an attractive feature if the model is used to characterize a nominal interest rate.[12]

[11] This process is a specific case of the more general *constant elasticity of variance process* given by $dx = (a_0 + a_1 x)\, dt + b_0 x^c dq$ where $c \in [0, 1]$.

[12] In contrast, recall from Chapter 9 that the Vasicek model (Vasicek 1977) assumes that the risk-free rate follows an Ornstein-Uhlenbeck process, which implies that r has a discrete-time normal distribution. Hence, the Vasicek model may be preferred for modeling a real interest rate since r can become negative. See (Pennacchi 1991) for such an application. It can be shown that the discrete-time distribution for the CIR interest rate process in (13.49) is a

Next, let us consider how to value contingent claims based on this example's assumptions. Specifically, let us consider the price of a default-free discount bond that pays one unit of the consumption good when it matures at date $T \geq t$. Since this bond's payoff is independent of wealth, and since logarithmic utility implies that the equilibrium interest rate and optimal portfolio proportions are independent of wealth, the price of this bond will also be independent of wealth. Hence, the derivatives P_W, P_{WW}, and P_{Wx} in the valuation equation (13.42) will all be zero. Moreover, since $r = \alpha x$, it will be insightful to think of r as the state variable rather than x, so that the date t bond price can be written as $P(r,t,T)$. With these changes, the valuation equation (13.42) becomes[13]

$$\frac{\sigma^2 r}{2} P_{rr} + \left[\kappa\left(\overline{r} - r\right) - \psi r\right] P_r - rP + P_t = 0 \tag{13.50}$$

where ψ is a constant equal to $\widehat{\boldsymbol{\omega}}'\widehat{\boldsymbol{\phi}}$. $\widehat{\boldsymbol{\omega}}$ equals the right-hand side of equation (13.45) but with $\boldsymbol{\mu}$ replaced by $\widehat{\boldsymbol{\mu}}$ and $\boldsymbol{\Omega}$ replaced by $\widehat{\boldsymbol{\Omega}}$, while $\widehat{\boldsymbol{\phi}}$ is an $n \times 1$ vector of constants whose i^{th} element is $\sigma \widehat{\sigma}_i \rho_i$. $\psi r = \boldsymbol{\omega}^{*\prime}\boldsymbol{\phi}$ is the covariance of interest rate changes with the proportional change in optimally invested wealth. In other words, it is the interest rate's "beta" (covariance with the market portfolio's return).

The partial differential equation (13.50), when solved subject to the boundary condition $P(r,T,T) = 1$, leads to the bond pricing formula

$$P(r,t,T) = A(\tau)\, e^{-B(\tau) r} \tag{13.51}$$

noncentral chi-square.

[13] Recall that logarithmic utility implies $A = W$ and $H = 0$.

where $\tau = T - t$,

$$A(\tau) \equiv \left[\frac{2\theta e^{(\theta+\kappa+\psi)\frac{\tau}{2}}}{(\theta+\kappa+\psi)(e^{\theta\tau}-1)+2\theta} \right]^{2\kappa\overline{r}/\sigma^2} \tag{13.52}$$

$$B(\tau) \equiv \frac{2\left(e^{\theta\tau}-1\right)}{(\theta+\kappa+\psi)(e^{\theta\tau}-1)+2\theta} \tag{13.53}$$

and $\theta \equiv \sqrt{(\kappa+\psi)^2 + 2\sigma^2}$. This CIR bond price can be contrasted with that of the Vasicek model derived in Chapter 9, equation (9.39). They are similar in having the same structure given in equation (13.51) but with different values for $A(\tau)$ and $B(\tau)$. Hence, the discount bond yield, $Y(r,\tau) \equiv -\ln[P(r,t,T)]/\tau = -\ln[A(\tau)]/\tau + B(\tau) r/\tau$, is linear in the state variable for both models.[14] But the two models differ in a number of ways. Recall that Vasicek directly assumed that the short rate, r, followed an Ornstein-Uhlenbeck process and derived the result that, in the absence of arbitrage, the market price of interest rate risk must be the same for bonds of all maturities. Using the notation of $\mu_p(r,\tau)$ and $\sigma_p(\tau)$ to be the mean and standard deviation of the return on a bond with τ periods to maturity, it was assumed that the market price of interest rate risk, $\left[\mu_p(r,\tau) - r\right]/\sigma_p(\tau)$, was a constant.

In contrast, the CIR model derived an equilibrium square root process for r based on assumptions of economic fundamentals (tastes and technologies). Moreover, the derivation of bond prices did not focus on the absence of arbitrage but rather the (zero-net-supply) market clearing conditions consistent with individuals' consumption and portfolio choices. Moreover, unlike the Vasicek model, the CIR derivation required no explicit assumption regarding the form of the market price of interest rate risk. Rather, this market price of risk was endogenous to the model's other assumptions regarding preferences and

[14] Models having bond yields that are linear in the state variables are referred to as *affine* models of the term structure. Such models will be discussed further in Chapter 17.

technologies. Let's solve for the market risk premium implicit in CIR bond prices.

Note that Itô's lemma says that the bond price follows the process

$$\begin{aligned} dP &= P_r dr + \frac{1}{2} P_{rr} \sigma^2 r dt + P_t dt \\ &= \left(\frac{1}{2} P_{rr} \sigma^2 r + P_r \left[\kappa \left(\overline{r} - r \right) \right] + P_t \right) dt + P_r \sigma \sqrt{r} d\zeta \end{aligned} \tag{13.54}$$

In addition, rearranging (13.50) implies that $\frac{1}{2} P_{rr} \sigma^2 r + P_r \left[\kappa \left(\overline{r} - r \right) \right] + P_t = rP + \psi r P_r$. Substituting this into (13.54), it can be rewritten as

$$\begin{aligned} dP/P &= r \left(1 + \psi \frac{P_r}{P} \right) dt + \frac{P_r}{P} \sigma \sqrt{r} d\zeta \\ &= r \left(1 - \psi B \left(\tau \right) \right) dt - B \left(\tau \right) \sigma \sqrt{r} d\zeta \end{aligned} \tag{13.55}$$

where we have used equation (13.51)'s result that $P_r / P = -B(\tau)$ in the second line of (13.55). Hence, we can write

$$\frac{\mu_p \left(r, \tau \right) - r}{\sigma_p \left(r, \tau \right)} = \frac{-\psi r B \left(\tau \right)}{\sigma \sqrt{r} B \left(\tau \right)} = -\frac{\psi \sqrt{r}}{\sigma} \tag{13.56}$$

so that the market price of interest rate risk is not constant, as in the Vasicek model, but is proportional to the square root of the interest rate. When $\psi < 0$, which occurs when the interest rate is negatively correlated with the return on the market portfolio (and bond prices are positively correlated with the market portfolio), bonds will carry a positive risk premium. CIR (Cox, Ingersoll, and Ross 1985b) argue that their equilibrium approach to deriving a market risk premium avoids problems that can occur when, following the no-arbitrage approach, an arbitrary form for a market risk premium is assumed. They show that some functional forms for market risk premia are inconsistent with the no-arbitrage assumption.

13.4 Summary

In a multiperiod, continuous-time environment, the Merton ICAPM shows that when investment opportunities are constant, the expected returns on assets satisfy the single-period CAPM relationship. For the more interesting case of changing investment opportunities, the CAPM relationship is generalized to include risk premia reflecting an asset's covariances with asset portfolios that best hedge against changes in investment opportunities. However, this multibeta relationship can be simplified to express an asset's expected return in terms of a single consumption beta.

The Cox, Ingersoll, and Ross model of a production economy helps to justify the ICAPM results by showing that they are consistent with a model that starts from more primitive assumptions regarding the nature of an economy's asset supplies. It also can be used to derive the economy's equilibrium risk-free interest rate and the shadow prices of contingent claims that are assumed to be in zero net supply. One important application of the model is a derivation of the equilibrium term structure of interest rates.

The next chapter builds on our results to this point by generalizing individuals' lifetime utility functions. No longer will we assume that utility is time separable. Allowing for time-inseparable utility can lead to different equilibrium relationships between asset returns that can better describe empirical findings.

13.5 Exercises

1. Consider a CIR economy similar to the log utility example given in this chapter. However, instead of the productive technologies following the

processes of equation (13.46), assume that they satisfy

$$d\eta_i/\eta_i \;=\; \widehat{\mu}_i x\, dt \;+\; \sigma_i dz_i, \quad i = 1, ..., n$$

In addition, rather than assume that the state variable follows the process (13.47), suppose that it is given by

$$dx = (a_0 + a_1 x)\, dt + b_0 d\zeta$$

where $dz_i d\zeta = \rho_i dt$. It is assumed that $a_0 > 0$ and $a_1 < 0$.

a. Solve for the equilibrium risk-free interest rate, r, and the process it follows, dr. What parametric assumptions are needed for the unconditional mean of r to be positive?

b. Derive the optimal (market) portfolio weights for this economy, $\boldsymbol{\omega}^*$. How does $\boldsymbol{\omega}^*$ vary with r?

c. Derive the partial differential equation for $P(r, t, T)$, the date t price of a default-free discount bond that matures at date T. Does this equation look familiar?

2. Consider the intertemporal consumption-portfolio choice model and the Intertemporal Capital Asset Pricing Model of Merton and its general equilibrium specification by Cox, Ingersoll, and Ross.

a. What assumptions are needed for the single-period Sharpe-Treyner-Linter-Mossin CAPM results to hold in this multiperiod environment where consumption and portfolio choices are made continuously?

b. Briefly discuss the portfolio choice implications of a situation in which the instantaneous real interest rate, $r(t)$, is stochastic, following a mean-reverting process such as the square root process of Cox, Ingersoll, and Ross or the Ornstein-Uhlenbeck process of Vasicek. Specifically, suppose that individuals can hold the instantaneous-maturity risk-free asset, a long-maturity default-free bond, and equities (stocks) and that a rise in $r(t)$ raises all assets' expected rates of return. How would the results differ from the single-period Markowitz portfolio demands? In explaining your answer, discuss how the results are sensitive to utility displaying greater or lesser risk aversion compared to log utility.

3. Consider a continuous-time version of a Lucas endowment economy. Let C_t be the aggregate dividends paid at date t, which equals aggregate consumption at date t. It is assumed to follow the lognormal process

$$dC/C = \mu_c dt + \sigma_c dz_c \tag{1}$$

where μ_c and σ_c are constants. The economy is populated with representative individuals whose lifetime utility is of the form

$$E_t \left[\int_t^\infty e^{-\rho s} \frac{C_s^\gamma}{\gamma} ds \right] \tag{2}$$

a. Solve for the process followed by the continuous-time pricing kernel, M_t. In particular, relate the equilibrium instantaneous risk-free interest rate and the market price of risk to the parameters in equation (1) and utility function (2) above.

b. Suppose that a particular risky asset's price follows the process

$$dS/S = \mu_s dt + \sigma_s dz_s$$

where $dz_s dz_c = \rho_{sc} dt$. Derive a value for μ_s using the pricing kernel process.

c. From the previous results, show that Merton's Intertemporal Capital Asset Pricing Model (ICAPM) and Breeden's Consumption Capital Asset Pricing Model (CCAPM) hold between this particular risky asset and the market portfolio of all risky assets.

Chapter 14

Time-Inseparable Utility

In previous chapters, individuals' multiperiod utility functions were assumed to be time separable. In a continuous-time context, time-separable expected lifetime utility was specified as

$$E_t\left[\int_t^T U\left(C_s,\, s\right)\, ds\right] \tag{14.1}$$

where $U\left(C_s,\, s\right)$ is commonly taken to be of the form

$$U\left(C_s,\, s\right) = e^{-\rho(s-t)}u\left(C_s\right) \tag{14.2}$$

so that utility at date s depends only on consumption at date s and not consumption at previous or future dates. However, as was noted earlier, there is substantial evidence that standard time-separable utility appears inconsistent with the empirical time series properties of U.S. consumption data and the average returns on risky assets (common stocks) and risk-free investments. These empirical contradictions, referred to as the equity premium puzzle and the risk-free interest rate puzzle, have led researchers to explore lifetime utility functions that differ from function (14.1) by permitting more general time-inseparable forms.

In this chapter we consider two types of lifetime utility functions that are not time separable. The first type is a class of lifetime utility functions for which *past* consumption plays a role in determining current utility. These utility functions display *habit persistence.* We summarize two models of this type, one by George Constantinides (Constantinides 1990) and the other by John Campbell and John Cochrane (Campbell and Cochrane 1999). In addition to modeling habit persistence differently, these models provide interesting contrasts in terms of their assumptions regarding the economy's aggregate supplies of assets and the techniques we can use to solve them. Constantinides' *internal* habit persistence model is a simple example of a Cox, Ingersoll, and Ross production economy (Cox, Ingersoll, and Ross 1985a) where asset supplies are perfectly elastic. It is solved using a Bellman equation approach. Campbell and Cochrane present a model of *external* habit persistence or "Keeping Up with the Joneses" preferences. Their model assumes a Lucas endowment economy (Lucas 1978) where asset supplies are perfectly inelastic. Its solution is based on the economy's stochastic discount factor.

The second type of time-inseparable utility that we discuss is called *recursive utility.* From one perspective, recursive utility is the opposite of habit persistence because recursive utility functions make current utility depend on expected values of *future* utility, which in turn depends on *future* consumption. We illustrate this type of utility by considering the general equilibrium of an economy where representative consumer-investors have recursive utility. The specific model that we analyze is a continuous-time version of a discrete-time model by Maurice Obstfeld (Obstfeld 1994). A useful aspect of this model is that it enables us to easily distinguish between an individual's coefficient of relative risk aversion and his elasticity of intertemporal substitution.

By generalizing utility functions to permit habit persistence or to be recursive, we hope to provide better models of individuals' actual preferences and their resulting consumption and portfolio choice decisions. In this way, greater insights into the nature of equilibrium asset returns may be possible. Specifically, we can analyze these models in terms of their ability to resolve various asset pricing "puzzles," such as the equity premium puzzle and the risk-free rate puzzle that arise when utility is time separable. Let us first investigate how utility can be extended from the standard time-separable, constant relative-risk-aversion case to display habit persistence. We then follow this with an examination of recursive utility.

14.1 Constantinides' Internal Habit Model

The notion of habit persistence can be traced to the writings of Alfred Marshall (Marshall 1920), James Duesenberry (Duesenberry 1949), and more recently, Harl Ryder and Geoffrey Heal (Ryder and Heal 1973). It is based on the idea that an individual's choice of consumption affects not only utility today but directly affects utility in the near future because the individual becomes accustomed to today's consumption standard.

Let us illustrate this idea by presenting Constantinides' internal habit formation model, which derives a representative individual's consumption and portfolio choices in a simple production economy. It is based on the following assumptions.

14.1.1 Assumptions

Technology

A single capital-consumption good can be invested in up to two different technologies. The first is a risk-free technology whose output, B_t, follows the process

$$dB/B \;=\; r\,dt \tag{14.3}$$

The second is a risky technology whose output, η_t, follows the process

$$d\eta/\eta \;=\; \mu\,dt \;+\; \sigma\,dz \tag{14.4}$$

Note that the specification of technologies fixes the expected rates of return and variances of the safe and risky investments.[1] In this setting, individuals' asset demands determine equilibrium quantities of the assets supplied rather than asset prices. Since r, μ, and σ are assumed to be constants, there is a constant investment opportunity set.

Preferences

Representative agents maximize expected utility of consumption, C_t, of the form

$$E_0\left[\int_0^\infty e^{-\rho t}u\left(\widehat{C}_t\right)\,dt\right] \tag{14.5}$$

where $u\left(\widehat{C}_t\right) = \widehat{C}_t^{\gamma}/\gamma$, $\gamma < 1$, $\widehat{C}_t = C_t - bx_t$, and

$$x_t \;\equiv\; e^{-at}x_0 \;+\; \int_0^t e^{-a(t-s)}C_s\,ds \tag{14.6}$$

Note that if $b = 0$, utility is of the standard time-separable form and displays constant relative risk aversion with a coefficient of relative risk aversion equal to $(1-\gamma)$. The variable x_t is an exponentially weighted sum of past consumption, so that when $b > 0$, the quantity bx_t can be interpreted as a "subsistence,"

[1] In this model, the existence of a risk-free technology determines the risk-free interest rate. This differs from our earlier presentation of the Cox, Ingersoll, and Ross model (Cox, Ingersoll, and Ross 1985a) where risk-free borrowing and lending is assumed to be in zero net supply and the interest rate is an equilibrium rate determined by risky investment opportunities and individuals' preferences.

or "habit," level of consumption and $\widehat{C}_t = C_t - bx_t$ can be interpreted as "surplus" consumption. In this case, the specification in (14.5) assumes that the individual's utility depends on only the level of consumption in excess of the habit level. This models the notion that an individual becomes accustomed to a standard of living (habit), and current utility derives from only the part of consumption that is in excess of this standard. Alternatively, if $b < 0$ so that past consumption adds to rather than subtracts from current utility, then the model can be interpreted as one displaying *durability* in consumption rather than habit persistence.[2] Empirical evidence comparing habit formation versus durability in consumption is mixed.[3] Research that models utility as depending on the consumptions of multiple goods, where some goods display habit persistence and others display durability in consumption, may be a better approach to explaining asset returns.[4] However, for simplicity, here we assume the single-good, $b > 0$ case introduced by Constantinides.

The Constantinides model of habit persistence makes current utility depend on a linear combination of not only current consumption but of past consumption through the variable x_t. Hence, it is not time separable. An increase in consumption at date t decreases current marginal utility, but it also increases the marginal utility of consumption at future dates because it raises the level of subsistence consumption. Of course, there are more general ways of modeling habit persistence, for example, $u(C_t, w_t)$ where w_t is any function of past consumption levels.[5] However, the linear habit persistence specification in (14.5)

[2] Ayman Hindy and Chi-Fu Huang (Hindy and Huang 1993) consider such a model.

[3] Empirical asset pricing tests by Wayne Ferson and George Constantinides (Ferson and Constantinides 1991) that used seasonally adjusted aggregate consumption data provided more support for habit persistence relative to consumption durability. In contrast, John Heaton (Heaton 1995) found more support for durability after adjusting for time-averaged data and seasonality.

[4] Multiple-good models displaying durability and habit persistence and durability have been developed by Jerome Detemple, Christos Giannikos, and Zhihong Shi (Detemple and Giannikos 1996); (Giannikos and Shi 2006).

[5] Jerome Detemple and Fernando Zapatero (Detemple and Zapatero 1991) consider a model that displays nonlinear habit persistence.

and (14.6) is attractive due to its analytical tractability.

Additional Parametric Assumptions

Let W_0 be the initial wealth of the representative individual. The following parametric assumptions are made to have a well-specified consumption and portfolio choice problem.

$$W_0 \; > \; \frac{bx_0}{r+a-b} \; > \; 0 \tag{14.7}$$

$$r+a \; > \; b \; > \; 0 \tag{14.8}$$

$$\rho - \gamma r - \frac{\gamma(\mu-r)^2}{2(1-\gamma)\sigma^2} \; > \; 0 \tag{14.9}$$

$$0 \; \leq \; m \; \equiv \; \frac{\mu - r}{(1-\gamma)\sigma^2} \; \leq \; 1 \tag{14.10}$$

The reasons for making these parametric assumptions are the following. Note that C_t needs to be greater than bx_t for the individual to avoid infinite marginal utility.[6] Conditions (14.7) and (14.8) ensure that an admissible (feasible) consumption and portfolio choice strategy exists that enables $C_t > bx_t$.[7] To see this, note that the dynamics for the individual's wealth are given by

$$dW \;=\; \{[(\mu-r)\omega_t + r]W - C_t\}\; dt \;+\; \sigma\omega_t W\, dz \tag{14.11}$$

where ω_t, $0 \leq \omega_t \leq 1$ is the proportion of wealth that the individual invests in the risky technology. Now if $\omega_t = 0$ for all t, that is, one invests only in the riskless technology, and consumption equals a fixed proportion of wealth,

[6] Note that $\lim_{C_t \to bx_t} \; (C_t - bx_t)^{-(1-\gamma)} = \infty$.

[7] The ability to maintain $C_t > bx_t$ is possible when the underlying economy is assumed to be a production economy because individuals have the freedom of determining the aggregate level of consumption versus savings. This is not possible in an endowment economy where the path of C_t and, therefore, its exponentially weighted average, x_t, is assumed to be an exogenous stochastic process. For many random processes, there will be a positive probability that $C_t < bx_t$. Based on this observation, David Chapman (Chapman 1998) argues that many models that assume a linear habit persistence are incompatible with an endowment economy equilibrium.

$C_t = (r + a - b)W_t$, then

$$dW \;=\; \{rW - (r + a - b)W\}\; dt \;=\; (b - a)\, W dt \tag{14.12}$$

which is a first-order differential equation in W having the initial condition that it equal W_0 at $t = 0$. Its solution is

$$W_t \;=\; W_0 e^{(b-a)t} \;\; > 0 \tag{14.13}$$

so that wealth always stays positive. This implies $C_t = (r+a-b)\, W_0\, e^{(b-a)t} > 0$ and

$$\begin{aligned} C_t - bx_t \quad &= \quad (r + a - b)W_0 e^{(b-a)t} - b\left[e^{-at}x_0 + \int_0^t e^{-a(t-s)}(r + a - b)\, W_0\, e^{(b-a)s}\, ds\right] \\ &= \quad (r + a - b)\, W_0\, e^{(b-a)t} - \left[e^{-at}bx_0 + b(r + a - b)W_0 e^{-at}\int_0^t e^{bs}\, ds\right] \\ &= \quad (r + a - b)\, W_0\, e^{(b-a)t} - \left[\, e^{-at}bx_0 + (r + a - b)W_0 e^{-at}(e^{bt} - 1)\,\right] \\ &= \quad e^{-at}\left[\,(r + a - b)W_0 - bx_0\,\right] \end{aligned} \tag{14.14}$$

which is greater than zero by assumption (14.7).

Condition (14.9) is a transversality condition. It ensures that if the individual follows an optimal policy (which will be derived next), the expected utility of consumption over an infinite horizon is finite. As will be seen, condition (14.10) ensures that the individual chooses to invest a nonnegative amount of wealth in the risky and risk-free technologies, since short-selling physical investments is infeasible. Recall from Chapter 12, equation (12.35) that m is the optimal choice of the risky-asset portfolio weight for the time-separable, constant relative-risk-aversion case.

14.1.2 Consumption and Portfolio Choices

The solution technique presented here uses a dynamic programming approach similar to that of (Sundaresan 1989) and our previous derivation of consumption and portfolio choices under time-separable utility.[8] The individual's maximization problem is

$$\max_{\{C_s,\,\omega_s\}} E_t\left[\int_t^\infty e^{-\rho\, s}\frac{[C_s - bx_s]^\gamma}{\gamma}\,ds\right] \;\equiv\; e^{-\rho\, t}J(W_t,\,x_t) \tag{14.15}$$

subject to the intertemporal budget constraint given by equation (14.11). Given the assumption of an infinite horizon, we can simplify the analysis by separating out the factor of the indirect utility function that depends on calendar time, t; that is, $\widehat{J}\,(W_t, x_t, t) = e^{-\rho\, t}J(W_t,\,x_t)$. The "discounted" indirect utility function depends on two state variables: wealth, W_t, and the state variable x_t, the current habit level of consumption. Since there are no changes in investment opportunities (μ, σ, and r are all constant), there are no other relevant state variables. Similar to wealth, x_t is not exogenous but depends on past consumption. We can work out its dynamics by taking the derivative of equation (14.6):

$$dx/dt \;=\; -ae^{-at}x_0 \;+\; C_t \;-\; a\int_0^t e^{-a(t-s)}C_s\,ds, \qquad \text{or} \tag{14.16}$$

$$dx \;=\; (C_t - ax_t)\;dt \tag{14.17}$$

[8] Interestingly, Mark Schroder and Costis Skiadas (Schroder and Skiadas 2002) show that consumption-portfolio choice models where an individual displays linear habit formation can be transformed into a consumption-portfolio model where the individual does not exhibit habit formation. This can often simplify solving such problems. Further, known solutions to time-separable or recursive utility consumption-portfolio choice problems can be transformed to obtain novel solutions that also display linear habit formation.

Thus, changes in x_t are instantaneously deterministic. The Bellman equation is then

$$\begin{aligned} 0 &= \max_{\{C_t,\omega_t\}} \left\{ U(C_t,\, x_t,\, t) + L[e^{-\rho t}J] \right\} \\ &= \max_{\{C_t,\omega_t\}} \left\{ e^{-\rho t}\gamma^{-1}(C_t - bx_t)^{\gamma} + e^{-\rho t}J_W[((\mu - r)\omega_t + r)W - C_t] \right. \\ &\quad \left. + \frac{1}{2}e^{-\rho t}J_{WW}\sigma^2\omega_t^2W^2 \ + \ e^{-\rho t}J_x\,(C_t - ax_t) - \rho e^{-\rho t}J \right\} \end{aligned} \tag{14.18}$$

The first-order conditions with respect to C_t and ω_t are

$$\begin{aligned} (C_t - bx_t)^{\gamma-1} &= J_W - J_x, \quad \text{or} \\ C_t &= bx_t + [J_W - J_x]^{\frac{1}{\gamma-1}} \end{aligned} \tag{14.19}$$

and

$$\begin{aligned} (\mu - r)WJ_W + \omega_t\sigma^2W^2J_{WW} &= 0, \quad \text{or} \\ \omega_t &= -\frac{J_W}{J_{WW}W}\,\frac{\mu - r}{\sigma^2} \end{aligned} \tag{14.20}$$

Note that the additional term $-J_x$ in (14.19) reflects the fact that an increase in current consumption has the negative effect of raising the level of subsistence consumption, which decreases future utility. The form of (14.20), which determines the portfolio weight of the risky asset, bears the same relationship to indirect utility as in the time-separable case.

Substituting (14.19) and (14.20) back into (14.18), we obtain the equilibrium partial differential equation:

$$\frac{1-\gamma}{\gamma}\left[J_W - J_x\right]^{\frac{-\gamma}{1-\gamma}} - \frac{J_W^2}{J_{WW}}\frac{(\mu-r)^2}{2\sigma^2} + (rW - bx)J_W + (b-a)xJ_x - \rho J \;=\; 0 \tag{14.21}$$

From our previous discussion of the time-separable, constant relative-risk-aversion case ($a = b = x = 0$), when the horizon is infinite, we saw from (12.33) that a solution for J is of the form $J(W) = kW^\gamma$. For this previous case, $u = C^\gamma/\gamma$, $u_c = J_W$, and optimal consumption was a constant proportion of wealth:

$$C_t^* \;=\; (\gamma k)^{\frac{1}{(\gamma-1)}} W_t \;=\; W_t\left[\rho - r\gamma - \frac{1}{2}(\frac{\gamma}{1-\gamma})\frac{(\mu-r)^2}{\sigma^2}\right] / (1-\gamma) \tag{14.22}$$

and

$$\omega_t^* \;=\; m \tag{14.23}$$

where m is defined in condition (14.10).

These results for the time-separable case suggest that the derived utility-of-wealth function for the time-inseparable case might have the form

$$J(W,\, x) \;=\; k_0[W + k_1 x]^\gamma \tag{14.24}$$

Making this guess, substituting it into (14.21), and setting the coefficients on x and W equal to zero, we find

$$k_0 \;=\; \frac{(r+a-b)h^{\gamma-1}}{(r+a)\gamma} \tag{14.25}$$

where

$$h \;\equiv\; \frac{r+a-b}{(r+a)(1-\gamma)}\left[\rho - \gamma r - \frac{\gamma(\mu-r)^2}{2(1-\gamma)\sigma^2}\right] \;>\; 0 \tag{14.26}$$

and

$$k_1 \;=\; -\frac{b}{r+a-b} < 0. \tag{14.27}$$

Using equations (14.19) and (14.20), this implies

$$C_t{}^* = bx_t + h\left[W_t - \frac{bx_t}{r+a-b}\right] \tag{14.28}$$

and

$$\omega_t^* = m\left[1 - \frac{bx_t/W_t}{r+a-b}\right] \tag{14.29}$$

Interestingly, since $r + a > b$, by assumption, the individual always demands less of the risky asset compared to the case of no habit persistence. Thus we would expect lower volatility of wealth over time.

In order to study the dynamics of C_t^*, consider the change in the term $\left[W_t - \frac{bx_t}{r+a-b}\right]$. Recall that the dynamics of W_t and x_t are given in equations (14.11) and (14.17), respectively. Using these, one finds

$$d\left[W_t - \frac{bx_t}{r+a-b}\right] = \left\{[\,(\mu - r)\omega_t^* + r]W_t - C_t^* - b\frac{C_t^* - ax_t}{r+a-b}\right\}dt + \sigma\omega_t^* W_t\, dz \tag{14.30}$$

Substituting in for ω_t^* and C_t^* from (14.28) and (14.29), one obtains

$$d\left[W_t - \frac{bx_t}{r+a-b}\right] = \left[W_t - \frac{bx_t}{r+a-b}\right][\,n\, dt + m\sigma\, dz] \tag{14.31}$$

where

$$n \equiv \frac{r-\rho}{1-\gamma} + \frac{(\mu - r)^2(2-\gamma)}{2(1-\gamma)^2\sigma^2} \tag{14.32}$$

Using this and (14.28), one can show[9]

$$\frac{dC_t}{C_t} = \left[n + b - \frac{(n+a)bx_t}{C_t}\right]dt + \left(\frac{C_t - bx_t}{C_t}\right)m\sigma\, dz \tag{14.33}$$

For particular parametric conditions, the ratio $\frac{bx_t}{C_t - bx_t}$ has a stationary distribu-

[9] See Appendix A in (Constantinides 1990).

tion.[10] However, one sees from the stochastic term in (14.33), $\left(\frac{C_t - bx_t}{C_t}\right) m\sigma\, dz$, that consumption growth is smoother than in the case of no habit persistence. For a given equity (risky-asset) risk premium, this can imply relatively smooth consumption paths, even though risk aversion, γ, may not be of a very high magnitude. To see this, recall from Chapter 4's inequality (4.32) that the Hansen-Jagannathan (H-J) bound for the time-separable case can be written as

$$\left|\frac{\mu - r}{\sigma}\right| \leq (1-\gamma)\,\sigma_c \tag{14.34}$$

In the current case of habit persistence, from (14.33) we see that the instantaneous standard deviation of consumption growth is

$$\begin{aligned} \sigma_{c,t} &= \left(\frac{C_t - bx_t}{C_t}\right) m\sigma \\ &= \left(\frac{\widehat{C}_t}{C_t}\right)\left[\frac{\mu - r}{(1-\gamma)\,\sigma^2}\right]\sigma \end{aligned} \tag{14.35}$$

where, recall, that $\widehat{C}_t \equiv C_t - bx_t$ is defined as surplus consumption. If we define $S_t \equiv \widehat{C}_t / C_t$ as the *surplus consumption ratio*, we can rearrange equation (14.35) to obtain

$$\frac{\mu - r}{\sigma} = \frac{(1-\gamma)\,\sigma_{c,t}}{S_t} \tag{14.36}$$

Since $S_t \equiv \frac{C_t - bx_t}{C_t}$ is less than 1, we see by comparing (14.36) to (14.34) that habit persistence may help reconcile the empirical violation of the H-J bound. With habit persistence, the lower demand for the risky asset, relative to the time-separable case, can result in a higher equilibrium excess return on the risky asset and, hence, may aid in explaining the "puzzle" of a large equity premium. However, empirical work by Wayne Ferson and George Constantinides (Ferson and Constantinides 1991) that tests linear models of habit persistence suggests

[10] See Theorem 2 in (Constantinides 1990).

that these models cannot produce an equity risk premium as large as that found in historical equity returns.

Let us next turn to another approach to modeling habit persistence where an individual's habit level depends on the behavior of other individuals and, hence, is referred to as an *external habit*.

14.2 Campbell and Cochrane's External Habit Model

The Campbell-Cochrane external habit persistence model is based on the following assumptions.

14.2.1 Assumptions

Technology

Campbell and Cochrane consider a discrete-time endowment economy. Date t aggregate consumption, which also equals aggregate output, is denoted C_t, and it is assumed to follow an independent and identically distributed lognormal process:

$$\ln(C_{t+1}) - \ln(C_t) = g + \nu_{t+1} \tag{14.37}$$

where $v_{t+1} \sim N\left(0, \sigma^2\right)$.

Preferences

It is assumed that there is a representative individual who maximizes expected utility of the form

$$E_0\left[\sum_{t=0}^{\infty} \delta^t \frac{(C_t - X_t)^\gamma - 1}{\gamma}\right] \tag{14.38}$$

where $\gamma < 1$ and X_t denotes the "habit level." X_t is related to past consumption in the following nonlinear manner. Define the surplus consumption ratio, S_t, as

$$S_t \equiv \frac{C_t - X_t}{C_t} \tag{14.39}$$

Then the log of surplus consumption is assumed to follow the autoregressive process[11]

$$\ln(S_{t+1}) = (1-\phi)\ln\left(\overline{S}\right) + \phi\ln(S_t) + \lambda(S_t)\,\nu_{t+1} \tag{14.40}$$

where $\lambda(S_t)$, the *sensitivity function*, measures the proportional change in the surplus consumption ratio resulting from a shock to output growth. It is assumed to take the form

$$\lambda(S_t) = \frac{1}{\overline{S}}\sqrt{1 - 2\left[\ln(S_t) - \ln\left(\overline{S}\right)\right]} - 1 \tag{14.41}$$

and

$$\overline{S} = \sigma\sqrt{\frac{1-\gamma}{1-\phi}} \tag{14.42}$$

The lifetime utility function in (14.38) looks somewhat similar to (14.5) of the Constantinides model. However, whereas Constantinides assumes that an individual's habit level depends on his or her own level of past consumption, Campbell and Cochrane assume that an individual's habit level depends on everyone else's current and past consumption. Thus, in the Constantinides model, the individual's choice of consumption, C_t, affects his future habit level, bx_s, for all $s > t$, and he takes this into account in terms of how it affects his

[11] This process is locally equivalent to $\ln(X_t) = \phi\ln(X_{t-1}) + \lambda\ln(C_t)$ or $\ln(X_t) = \lambda\sum_{i=0}^{\infty}\phi^i\ln(C_{t-i})$. The reason for the more complicated form in (14.40) is that it ensures that consumption is always above habit since S_t is always positive. This precludes infinite marginal utility.

expected utility when he chooses C_t. This type of habit formation is referred to as *internal* habit. In contrast, in the Campbell and Cochrane model, the individual's choice of consumption, C_t, does not affect her future habit level, X_s, for all $s \geq t$, so that she views X_t as exogenous when choosing C_t. This type of habit formation is referred to as *external* habit or "Keeping Up with the Joneses."[12] The external habit assumption simplifies the representative agent's decision making because habit becomes an exogenous state variable that depends on aggregate, not the individual's, consumption.

14.2.2 Equilibrium Asset Prices

Because habit is exogenous to the individual, the individual's marginal utility of consumption is

$$u_c\left(C_t, X_t\right) = \left(C_t - X_t\right)^{\gamma-1} = C_t^{\gamma-1} S_t^{\gamma-1} \tag{14.43}$$

and the representative agent's stochastic discount factor is

$$m_{t,t+1} = \delta \frac{u_c\left(C_{t+1}, X_{t+1}\right)}{u_c\left(C_t, X_t\right)} = \delta \left(\frac{C_{t+1}}{C_t}\right)^{\gamma-1} \left(\frac{S_{t+1}}{S_t}\right)^{\gamma-1} \tag{14.44}$$

If we define r as the continuously compounded, risk-free real interest rate between dates t and $t+1$, then it equals

$$\begin{aligned} r &= -\ln\left(E_t\left[m_{t,t+1}\right]\right) = -\ln\left(\delta E_t\left[e^{-(1-\gamma)\ln(C_{t+1}/C_t)-(1-\gamma)\ln(S_{t+1}/S_t)}\right]\right) \\ &= -\ln\left(\delta e^{-(1-\gamma)E_t[\ln(C_{t+1}/C_t)]-(1-\gamma)E_t[\ln(S_{t+1}/S_t)]+\frac{1}{2}(1-\gamma)^2 Var_t[\ln(C_{t+1}/C_t)+\ln(S_{t+1}/S_t)]}\right) \\ &= -\ln\left(\delta\right) + \left(1-\gamma\right) g + \left(1-\gamma\right)\left(1-\phi\right)\left(\ln \overline{S} - \ln S_t\right) - \frac{\left(1-\gamma\right)^2 \sigma^2}{2}\left[1+\lambda\left(S_t\right)\right]^2 \end{aligned} \tag{14.45}$$

[12] A similar modeling was developed by Andrew Abel (Abel 1990).

Substituting in for $\lambda(S_t)$ from (14.41), equation (14.45) becomes

$$r = -\ln(\delta) + (1-\gamma)\,g - \frac{1}{2}(1-\gamma)(1-\phi) \tag{14.46}$$

which, by construction, turns out to be constant over time. One can also derive a relationship for the date t price of the market portfolio of all assets, denoted P_t. Recall that since we have an endowment economy, aggregate consumption equals the economy's aggregate output, which equals the aggregate dividends paid by the market portfolio. Therefore,

$$P_t = E_t\left[m_{t,t+1}\left(C_{t+1} + P_{t+1}\right)\right] \tag{14.47}$$

or, equivalently, one can solve for the price-dividend ratio for the market portfolio:

$$\begin{aligned} \frac{P_t}{C_t} &= E_t\left[m_{t,t+1}\frac{C_{t+1}}{C_t}\left(1+\frac{P_{t+1}}{C_{t+1}}\right)\right] \qquad (14.48)\\ &= \delta E_t\left[\left(\frac{S_{t+1}}{S_t}\right)^{\gamma-1}\left(\frac{C_{t+1}}{C_t}\right)^{\gamma}\left(1+\frac{P_{t+1}}{C_{t+1}}\right)\right] \end{aligned}$$

As in the Lucas model, this stochastic difference equation can be solved forward to obtain

$$\begin{aligned} \frac{P_t}{C_t} &= \delta E_t\left[\left(\frac{S_{t+1}}{S_t}\right)^{\gamma-1}\left(\frac{C_{t+1}}{C_t}\right)^{\gamma}\left(1+\delta\left(\frac{S_{t+2}}{S_{t+1}}\right)^{\gamma-1}\left(\frac{C_{t+2}}{C_{t+1}}\right)^{\gamma}\left(1+\frac{P_{t+2}}{C_{t+2}}\right)\right)\right]\\ &= E_t\left[\delta\left(\frac{S_{t+1}}{S_t}\right)^{\gamma-1}\left(\frac{C_{t+1}}{C_t}\right)^{\gamma} + \delta^2\left(\frac{S_{t+2}}{S_t}\right)^{\gamma-1}\left(\frac{C_{t+2}}{C_t}\right)^{\gamma} + ...\right]\\ &= E_t\left[\sum_{i=1}^{\infty}\delta^i\left(\frac{S_{t+i}}{S_t}\right)^{\gamma-1}\left(\frac{C_{t+i}}{C_t}\right)^{\gamma}\right] \qquad (14.49) \end{aligned}$$

The solutions can then be computed numerically by simulating the lognormal processes for C_t and S_t. The distribution of C_{t+1}/C_t is lognormal and does not depend on the level of consumption, C_t, whereas the distribution of S_{t+1}/S_t does depend on the current level of S_t.[13] Hence, the value of the market portfolio relative to current output, P_t/C_t, varies only with the current surplus consumption ratio, S_t. By numerically calculating P_t/C_t as a function of S_t, Campbell and Cochrane can determine the market portfolio's expected returns and the standard deviation of returns as the level of S_t varies.

Note that in this model, the coefficient of relative risk aversion is given by

$$-\frac{C_t u_{cc}}{u_c} = \frac{1-\gamma}{S_t} \tag{14.50}$$

and, as was shown in inequality (4.32), the relationship between the Sharpe ratio for any asset and the coefficient of relative risk aversion when consumption is lognormally distributed is approximately

$$\left| \frac{E[r_i] - r}{\sigma_{r_i}} \right| \leq -\frac{C_t u_{cc}}{u_c} \sigma_c = \frac{(1-\gamma)\,\sigma_c}{S_t} \tag{14.51}$$

which has a similar form to that of the Constantinides internal habit model except, here, σ_c is a constant and, for the case of the market portfolio, $E[r_i]$ and σ_{r_i} will be time-varying functions of S_t. The coefficient of relative risk aversion will be relatively high when S_t is relatively low, that is, when consumption is low (a recession). Moreover, the model predicts that the equity risk premium increases during a recession (when $-\frac{C_t u_{cc}}{u_c}$ is high), a phenomenon that seems to be present in the postwar U.S. stock market. Campbell and Cochrane calibrate the model to U.S. consumption and stock market data.[14] Due to the different

---

[13] Note that from (14.37) expected consumption growth, g, is a constant, but from (14.40) the expected growth in the surplus consumption ratio, $(1-\phi)\left[\ln\left(\overline{S}\right) - \ln\left(S_t\right)\right]$, is mean-reverting.

[14] They generalize the model to allow dividends on the (stock) market portfolio to differ from consumption, so that dividend growth is not perfectly correlated with consumption growth. Technically, this violates the assumption of an endowment economy but, empirically, there is low correlation between growth rates of stock market dividends and consumption.

(nonlinear) specification for S_t vis-à-vis the model of Constantinides, they have relatively more success in fitting this model to data on asset prices.[15]

The next section introduces a class of time-inseparable utility that is much different from habit persistence in that current utility depends on expected future utility which, in turn, depends on future consumption. Hence, unlike habit persistence, in which utility depends on past consumption and is backward looking, recursive utility is forward looking.[16]

14.3 Recursive Utility

A class of time-inseparable utility known as recursive utility was developed by David Kreps and Evan Porteus (Kreps and Porteus 1978) and Larry Epstein and Stanley Zin (Epstein and Zin 1989). They analyze this type of utility in a discrete-time setting, while Darrell Duffie and Larry Epstein (Duffie and Epstein 1992a) study the continuous-time limit. In continuous time, recall that standard, time-separable utility can be written as

$$V_t = E_t \left[\int_t^T U\left(C_s, s\right) ds \right] \tag{14.52}$$

[15] Empirical tests of the Campbell-Cochrane model by Thomas Tallarini and Harold Zhang (Tallarini and Zhang 2005) confirm that the model fits variation in the equity risk premium over the business cycle. However, while the model matches the mean returns on stocks, it fails to match higher moments such as the variance and skewness of stock returns. Another study by Martin Lettau and Harald Uhlig (Lettau and Uhlig 2000) embeds Campbell and Cochrane's external habit preferences in a production economy model having a labor-leisure decision. In this environment, they find that individuals' consumption and labor market decisions are counterfactual to their actual business cycle dynamics.

[16] Utility can be both forward and backward looking in that it is possible to construct models that are recursive and also display habit persistence (Schroder and Skiadas 2002).

where $U(C_s, s)$ is often taken to be of the form $U(C_s, s) = e^{-\rho(s-t)}u(C_s)$. Recursive utility, however, is specified as

$$V_t = E_t\left[\int_t^T f(C_s, V_s)\, ds\right] \tag{14.53}$$

where f is known as an *aggregator function*. The specification is recursive in nature because current lifetime utility, V_t, depends on expected values of future lifetime utility, V_s, $s > t$. When f has appropriate properties, Darrell Duffie and Larry Epstein (Duffie and Epstein 1992b) show that a Bellman-type equation can be derived that characterizes the optimal consumption and portfolio choice policies for utility of this type. For particular functional forms, they have been able to work out a number of asset pricing models.

In the example to follow, we consider a form of recursive utility that is a generalization of standard power (constant relative-risk-aversion) utility in that it separates an individual's risk aversion from her elasticity of intertemporal substitution. This generalization is potentially important because, as was shown in Chapter 4, equation (4.14), multiperiod power utility restricts the elasticity of intertemporal substitution, ϵ, to equal $1/(1-\gamma)$, the reciprocal of the coefficient of relative risk aversion. Conceptually, this may be a strong restriction. Risk aversion characterizes an individual's (portfolio) choices between assets of different risks and is a well-defined concept even in an atemporal (single-period) setting, as was illustrated in Chapter 1. In contrast, the elasticity of intertemporal substitution characterizes an individual's choice of consumption at different points in time and is inherently a temporal concept.

14.3.1 A Model by Obstfeld

Let us now consider the general equilibrium of an economy where representative consumer-investors have recursive utility. We analyze the simple production economy model of Maurice Obstfeld (Obstfeld 1994). This model makes the following assumptions.

Technology

A single capital-consumption good can be invested in up to two different technologies. The first is a risk-free technology whose output, B_t, follows the process

$$dB/B = rdt \tag{14.54}$$

The second is a risky technology whose output, η_t, follows the process

$$d\eta/\eta = \mu dt + \sigma dz \tag{14.55}$$

As in the Constantinides model's production economy, the specification of technologies fixes the expected rates of return and variances of the safe and risky investments. Individuals' asset demands will determine equilibrium quantities of the assets supplied rather than asset prices. Since r, μ, and σ are assumed to be constants, there is a constant investment opportunity set.

Preferences

Representative, infinitely lived households must choose between consuming (at rate C_s at date s) and investing the single capital-consumption good in the two technologies. The lifetime utility function at date t faced by each of these households, denoted V_t, is

$$V_t = E_t \int_t^{\infty} f\left(C_s, V_s\right) ds \tag{14.56}$$

where f, the aggregator function, is given by

$$f(C_s, V_s) = \rho \frac{C_s^{1-\frac{1}{\epsilon}} - \left[\gamma V_s\right]^{\frac{\epsilon-1}{\epsilon\gamma}}}{\left(1-\frac{1}{\epsilon}\right)\left[\gamma V_s\right]^{\frac{\epsilon-1}{\epsilon\gamma}-1}} \tag{14.57}$$

Clearly, this specification is recursive in that current lifetime utility, V_t, depends on expected values of *future* lifetime utility, V_s, $s > t$. The form of equation (14.57) is ordinally equivalent to the continuous-time limit of the discrete-time utility function specified in (Obstfeld 1994). Recall that utility functions are ordinally equivalent; that is, they result in the same consumer choices, if the utility functions evaluated at equivalent sets of decisions produce values that are linear transformations of each other. It can be shown (see (Epstein and Zin 1989) and (Duffie and Epstein 1992a)) that $\rho > 0$ is the continuously compounded subjective rate of time preference; $\epsilon > 0$ is the household's elasticity of intertemporal substitution; and $1 - \gamma > 0$ is the household's coefficient of relative risk aversion. For the special case of $\epsilon = 1/(1-\gamma)$, the utility function given in (14.56) and (14.57) is (ordinally) equivalent to the time-separable, constant relative-risk-aversion case:

$$V_t = E_t \int_t^{\infty} e^{-\rho s} \frac{C_s^{\gamma}}{\gamma} ds \tag{14.58}$$

Let ω_t be the proportion of each household's wealth invested in the risky asset (technology). Then the intertemporal budget constraint is given by

$$dW = [\omega(\mu - r)W + rW - C]\, dt + \omega\sigma W dz \tag{14.59}$$

When the aggregator function, f, is put in a particular form by an ordinally equivalent change in variables, what Duffie and Epstein (Duffie and Epstein 1992b) refer to as a "normalization," then a Bellman equation can be used to solve the problem. The aggregator in (14.57) is in normalized form.

As before, let us define $J(W_t)$ as the maximized lifetime utility at date t:

$$\begin{aligned} J(W_t) &= \max_{\{C_s,\omega_s\}} E_t \int_t^\infty f(C_s, V_s)\, ds \\ &= \max_{\{C_s,\omega_s\}} E_t \int_t^\infty f(C_s, J(W_s))\, ds \end{aligned} \tag{14.60}$$

Since this is an infinite horizon problem with constant investment opportunities, and the aggregator function, $f(C,V)$, is not an explicit function of calendar time, the only state variable is W.

The solution to the individual's consumption and portfolio choice problem is given by the continuous-time stochastic Bellman equation

$$0 = \max_{\{C_t,\omega_t\}} f[C_t, J(W_t)] + L[J(W_t)] \tag{14.61}$$

or

$$\begin{aligned} 0 &= \max_{\{C_t,\omega_t\}} f[C, J(W)] + J_W[\omega(\mu - r)W + rW - C] + \frac{1}{2} J_{WW}\omega^2\sigma^2 W^2 \\ &= \max_{\{C_t,\omega_t\}} \rho \frac{C^{1-\frac{1}{\epsilon}} - [\gamma J]^{\frac{\epsilon-1}{\epsilon\gamma}}}{\left(1-\frac{1}{\epsilon}\right)[\gamma J]^{\frac{\epsilon-1}{\epsilon\gamma}-1}} + J_W[\omega(\mu - r)W + rW - C] + \frac{1}{2} J_{WW}\omega^2\sigma^2 W^2 \end{aligned} \tag{14.62}$$

Taking the first-order condition with respect to C,

$$\rho \frac{C^{-\frac{1}{\epsilon}}}{[\gamma J]^{\frac{\epsilon-1}{\epsilon\gamma}-1}} - J_W = 0 \tag{14.63}$$

or

$$C = \left(\frac{J_W}{\rho}\right)^{-\epsilon} [\gamma J]^{\frac{1-\epsilon}{\gamma}+\epsilon} \tag{14.64}$$

Taking the first-order condition with respect to ω,

$$J_W(\mu - r)W + J_{WW}\omega\sigma^2 W^2 = 0 \tag{14.65}$$

or

$$\omega = -\frac{J_W}{J_{WW}W}\frac{\mu - r}{\sigma^2} \tag{14.66}$$

Substituting the optimal values for C and ω given by (14.64) and (14.66) into the Bellman equation (14.62), we obtain the ordinary differential equation:

$$\rho\frac{\left(\frac{J_W}{\rho}\right)^{1-\epsilon}[\gamma J]^{(\epsilon-1)\left[1-\frac{\epsilon-1}{\epsilon\gamma}\right]} - [\gamma J]^{\frac{1-\epsilon}{\epsilon\gamma}}}{\left(1-\frac{1}{\epsilon}\right)[\gamma J]^{\frac{\epsilon-1}{\epsilon\gamma}-1}} \tag{14.67}$$

$$+J_W\left[-\frac{J_W}{J_{WW}}\frac{(\mu-r)^2}{\sigma^2} + rW - \left(\frac{J_W}{\rho}\right)^{-\epsilon}[\gamma J]^{\frac{1-\epsilon}{\gamma}+\epsilon}\right] + \frac{1}{2}\frac{J_W^2}{J_{WW}}\frac{(\mu-r)^2}{\sigma^2} = 0$$

or

$$\frac{\epsilon\rho}{\epsilon-1}\left[\left(\frac{J_W}{\rho}\right)^{-\epsilon}[\gamma J]^{\frac{1-\epsilon}{\gamma}+\epsilon} - \gamma J\right] \tag{14.68}$$

$$+J_W\left[-\frac{J_W}{J_{WW}}\frac{(\mu-r)^2}{\sigma^2} + rW - \left(\frac{J_W}{\rho}\right)^{-\epsilon}[\gamma J]^{\frac{1-\epsilon}{\gamma}+\epsilon}\right] + \frac{1}{2}\frac{J_W^2}{J_{WW}}\frac{(\mu-r)^2}{\sigma^2} = 0$$

If one "guesses" that the solution is of the form $J(W) = (aW)^\gamma/\gamma$ and substitutes this into (14.68), one finds that $a = \alpha^{1/(1-\epsilon)}$ where

$$\alpha \equiv \rho^{-\epsilon}\left(\epsilon\rho + (1-\epsilon)\left[r + \frac{(\mu-r)^2}{2(1-\gamma)\sigma^2}\right]\right) \tag{14.69}$$

Thus, substituting this value for J into (14.64), we find that optimal consumption is a fixed proportion of wealth:

$$\begin{aligned} C &= \alpha\rho^{\epsilon}W \\ &= \left(\epsilon\rho + (1-\epsilon)\left[r + \frac{(\mu - r)^2}{2(1-\gamma)\sigma^2}\right]\right)W \end{aligned} \tag{14.70}$$

and the optimal portfolio weight of the risky asset is

$$\omega = \frac{\mu - r}{(1-\gamma)\sigma^2} \tag{14.71}$$

which is the same as for an individual with standard constant relative risk aversion and time-separable utility. The result that the optimal portfolio choice depends only on risk aversion turns out to be an artifact of the model's assumption that investment opportunities are constant. Harjoat Bhamra and Raman Uppal (Bhamra and Uppal 2003) demonstrate that when investment opportunities are stochastic, the portfolio weight, ω, can depend on both γ and ϵ.

Note that if $\epsilon = 1/(1-\gamma)$, then equation (14.70) is the same as optimal consumption for the time-separable, constant relative-risk-aversion, infinite horizon case given in Chapter 12, equation (12.34), $C = \frac{\gamma}{1-\gamma}\left[\frac{\rho}{\gamma} - r - \frac{(\mu-r)^2}{2(1-\gamma)\sigma^2}\right]W$. Similar to the time-separable case, for an infinite horizon solution to exist, we need consumption to be positive in (14.70), which requires $\rho > \frac{\epsilon-1}{\epsilon}\left(r + [\mu - r]^2 / \left[2(1-\gamma)\sigma^2\right]\right)$. This will be the case when the elasticity of intertemporal substitution, ϵ, is sufficiently small. For example, assuming $\rho > 0$, this inequality is always satisfied when $\epsilon < 1$.

14.3.2 Discussion of the Model

Let us examine how optimal consumption depends on the model's parameters. Note that the term $r + [\mu - r]^2 / \left[2(1-\gamma)\sigma^2\right]$ in (14.70) can be rewritten using $\omega = (\mu - r) / \left[(1-\gamma)\sigma^2\right]$ from (14.71) as

$$r + \frac{(\mu - r)^2}{2(1-\gamma)\sigma^2} = r + \omega\frac{\mu - r}{2} \tag{14.72}$$

and can be interpreted as relating to the risk-adjusted investment returns available to individuals. From (14.70) we see that an increase in (14.72) increases consumption when $\epsilon < 1$ and reduces consumption when $\epsilon > 1$. This result provides intuition for the role of intertemporal substitution. When $\epsilon < 1$, the income effect from an improvement in investment opportunities dominates the substitution effect, so that consumption rises and savings fall. The reverse occurs when $\epsilon > 1$: the substitution effect dominates the income effect and savings rise.

We can also study how the growth rate of the economy depends on the model's parameters. Assuming $0 < \omega < 1$ and substituting (14.70) and (14.71) into (14.59), we have that wealth follows the geometric Brownian motion process:

$$\begin{aligned} dW/W &= \left[\omega^* (\mu - r) + r - \alpha\rho^\epsilon\right] dt + \omega^* \sigma dz \qquad (14.73) \\ &= \left[\frac{(\mu - r)^2}{(1-\gamma)\sigma^2} + r - \epsilon\rho - (1-\epsilon)\left(r + \frac{(\mu - r)^2}{2(1-\gamma)\sigma^2}\right)\right] dt + \frac{\mu - r}{(1-\gamma)\sigma} dz \\ &= \left[\epsilon\left(r + \frac{(\mu - r)^2}{2(1-\gamma)\sigma^2} - \rho\right) + \frac{(\mu - r)^2}{2(1-\gamma)\sigma^2}\right] dt + \frac{\mu - r}{(1-\gamma)\sigma} dz \end{aligned}$$

Since $C = \alpha\rho^\epsilon W$, the drift and volatility of wealth in (14.73) are also the drift and volatility of the consumption process, dC/C. Thus, consumption and wealth are both lognormally distributed and their continuously compounded growth, $d \ln C$, has a volatility, σ_c, and mean, g_c, equal to

$$\sigma_c = \frac{\mu - r}{(1-\gamma)\,\sigma} \tag{14.74}$$

and

$$\begin{aligned} g_c &= \epsilon \left(r + \frac{(\mu - r)^2}{2(1-\gamma)\,\sigma^2} - \rho \right) + \frac{(\mu - r)^2}{2(1-\gamma)\,\sigma^2} - \frac{1}{2}\sigma_c^2 \\ &= \epsilon \left(r + \frac{(\mu - r)^2}{2(1-\gamma)\,\sigma^2} - \rho \right) - \frac{\gamma\,(\mu - r)^2}{2(1-\gamma)^2\,\sigma^2} \end{aligned} \tag{14.75}$$

From (14.75) we see that if $r + [\mu - r]^2 / \left[2(1-\gamma)\,\sigma^2\right] > \rho$, then an economy's growth rate is higher the higher is intertemporal substitution, ϵ, since individuals save more. Also, consider how an economy's rate varies with the squared Sharpe ratio, $[\mu - r]^2 / \sigma^2$, a measure of the relative attractiveness of the risky asset. The sign of the derivative $\partial g_c / \partial \left([\mu - r]^2 / \sigma^2 \right)$ equals the sign of $\epsilon - \gamma / (1-\gamma)$. For the time-separable, constant relative-risk-aversion case of $\epsilon = 1/(1-\gamma)$, this derivative is unambiguously positive, indicating that a higher μ or a lower σ would result in the economy growing faster. However, in the general case, the economy could grow slower if $\epsilon < \gamma/(1-\gamma)$. Why? Although from (14.71) we see that individuals put a large proportion of their wealth into the faster-growing risky asset as the Sharpe ratio rises, a higher Sharpe ratio leads to greater consumption (and less savings) when $\epsilon < 1$. For $\epsilon < \gamma/(1-\gamma)$, the effect of less savings dominates the portfolio effect and the economy is expected to grow more slowly.

Obstfeld points out that the integration of global financial markets that allows residents to hold risky foreign, as well as domestic, investments increases diversification and effectively reduces individuals' risky portfolio variance, σ^2.

This reduction in σ would lead individuals to allocate a greater proportion of their wealth to the higher-yielding risky assets. If $\epsilon > \gamma/(1-\gamma)$, financial market integration also would predict that countries would tend to grow faster.

It is natural to ask whether this recursive utility specification, which distinguishes between risk aversion and the intertemporal elasticity of substitution, can provide a better fit to historical asset returns compared to time-separable power utility. In terms of explaining the equity premium puzzle, from (14.74) we see that the risky-asset Sharpe ratio, $(\mu - r)/\sigma$, equals $(1-\gamma)\sigma_c$, the same form as with time-separable utility. So, as discussed earlier, one would still need to assume that the coefficient of relative risk aversion $(1-\gamma)$ were quite high in order to justify the equity risk premium. However, recursive utility has more hope of explaining the risk-free rate puzzle because of the additional degree of freedom added by the elasticity of substitution parameter, ϵ. If we substitute (14.74) into (14.75) and solve for the risk-free rate, we find

$$r = \rho + \frac{g_c}{\epsilon} - \left[1 - \gamma - \frac{\gamma}{\epsilon}\right]\frac{\sigma_c^2}{2} \tag{14.76}$$

Recall that for the time-separable case of $\epsilon = 1/(1-\gamma)$, we have

$$r = \rho + (1-\gamma)\, g_c - (1-\gamma)^2 \frac{\sigma_c^2}{2} \tag{14.77}$$

Because, empirically, $g_c \approx 0.018$ is large relative to $\sigma_c^2/2 \approx 0.03^2/2 = 0.00045$, the net effect of higher risk aversion, $1-\gamma$, needed to fit the equity risk premium leads to too high a risk-free rate in (14.77). However, we see that the recursive utility specification in (14.76) potentially circumvents this problem because g_c is divided by ϵ rather than being multiplied by $1-\gamma$.[17]

Empirical estimates of the elasticity of intertemporal substitution have been

[17] Philippe Weil (Weil 1989) appears to be the first to examine the equity premium and risk-free rate puzzles in the context of recursive utility.

obtained by regressing consumption growth, $d\ln C$, on the real interest rate, r. From equations (14.73) and (14.75), we see that if the risky-asset Sharpe ratio, $(\mu - r)/\sigma$, is assumed to be independent of the level of the real interest rate, r, then the regression coefficient on the real interest should provide an estimate of ϵ. Tests using aggregate consumption data, such as (Hall 1988) and (Campbell and Mankiw 1989), generally find that ϵ is small, often indistinguishable from zero. However, other tests based on consumption data disaggregated at the state level (Beaudry and van Wincoop 1996) or at the household level (Attanasio and Weber 1993) find higher estimates for ϵ, often around 1. From (14.76) we see that since $\sigma_c^2/2$ is small and assuming ρ is also small, a value of $\epsilon = 1$ could produce a reasonable value for the real interest rate.

14.4 Summary

The models presented in this chapter generalize the standard model of time-separable, power utility. For particular functional forms, an individual's consumption and portfolio choice problem can be solved using the same techniques that were previously applied to the time-separable case. For utility that displays habit persistence, we saw that the standard coefficient of relative risk aversion, $(1 - \gamma)$, is transformed to the expression $(1 - \gamma)/S_t$ where $S_t < 1$ is the surplus consumption ratio. Hence, habit persistence can make individuals behave in a very risk-averse fashion in order to avoid consuming below their habit or subsistence level. As a result, these models have the potential to produce aversion to holding risky assets sufficient to justify a high equity risk premium.

An attraction of recursive utility is that it distinguishes between an individual's level of risk aversion and his elasticity of intertemporal substitution, a distinction that is not possible with time-separable, power utility, which makes these characteristics reciprocals of one another. As a result, recursive utility can permit an individual to have high risk aversion while, at the same time, having a high elasticity of intertemporal substitution. Such a utility specification has the potential to produce both a high equity risk premium and a low risk-free interest rate that is present in historical data.

While recursive utility and utility displaying habit persistence might be considered nonstandard forms of utility, they are preference specifications that are considered to be those of rational individuals. In the next chapter we study utility that is influenced by psychological biases that might be described as irrational behavior. Such biases have been identified in experimental settings but have also been shown to be present in the actual investment behavior of some individuals. We examine how these biases might influence the equilibrium prices of assets.

14.5 Exercises

1. In the Constantinides habit persistence model, suppose that there are three, rather than two, technologies. Assume that there are the risk-free technology and two risky technologies:

$$\begin{aligned} dB/B &= rdt \\ dS_1/S_1 &= \mu_1 dt + \sigma_1 dz_1 \\ dS_2/S_2 &= \mu_2 dt + \sigma_2 dz_2 \end{aligned}$$

   where $dz_1 dz_2 = \phi dt$. Also assume that the parameters are such that there is an interior solution for the portfolio weights (all portfolio weights are positive). What would be the optimal consumption and portfolio weights for this case?

2. Consider an endowment economy where a representative agent maximizes utility of the form

$$\max \sum_{t=0}^{\infty} \delta^t \frac{(C_t - X_t)^\gamma}{\gamma}$$

where X_t is a level of external habit and equals $X_t = \theta \overline{C}_{t-1}$, where $\overline{C}_{t-1}$ is aggregate consumption at date $t-1$.

a. Write down an expression for the one-period, risk-free interest rate at date t, $R_{f,t}$.

b. If consumption growth, C_{t+1}/C_t, follows an independent and identical distribution, is the one-period riskless interest rate, $R_{f,t}$, constant over time?

3. The following problem is based on the work of Menzly, Santos, and Veronesi (Menzly, Santos, and Veronesi 2001). Consider a continuous-time endowment economy where agents maximize utility that displays external habit persistence. Utility is of the form

$$E_t \left[\int_0^\infty e^{-\rho t} \ln (C_t - X_t)\, dt \right]$$

and aggregate consumption (dividend output) follows the lognormal process

$$dC_t/C_t = \mu dt + \sigma dz$$

Define Y_t as the inverse surplus consumption ratio, that is, $Y_t \equiv \frac{C_t}{C_t - X_t}$ $= \frac{1}{1-(X_t/C_t)} > 1$. It is assumed to satisfy the mean-reverting process

$$dY_t = k\left(\overline{Y} - Y_t\right) dt - \alpha \left(Y_t - \lambda\right) dz$$

where $\overline{Y} > \lambda \geq 1$ is the long-run mean of the inverse surplus, $k > 0$ reflects the speed of mean reversion, $\alpha > 0$. The parameter λ sets a lower bound for Y_t, and the positivity of $\alpha (Y_t - \lambda)$ implies that a shock to the aggregate output (dividend-consumption) process decreases the inverse surplus consumption ratio (and increases the surplus consumption ratio). Let P_t be the price of the market portfolio. Derive a closed-form expression for the price-dividend ratio of the market portfolio, P_t/C_t. How does P_t/C_t vary with an increase in the surplus consumption ratio?

4. Consider an individual's consumption and portfolio choice problem when her preferences display habit persistence. The individual's lifetime utility satisfies

$$E_t \left[\int_t^T e^{-\rho s} u \left(C_s, x_s \right) \, ds \right] \tag{1}$$

where C_s is date s consumption and x_s is the individual's date s level of habit. The individual can choose among a risk-free asset that pays a constant rate of return equal to r and n risky assets. The instantaneous rate of return on risky asset i satisfies

$$dP_i/P_i = \mu_i dt + \sigma_i dz_i, \quad i = 1, ..., n \tag{2}$$

where $dz_i dz_j = \sigma_{ij} dt$ and μ_i, σ_i, and σ_{ij} are constants. Thus, the individual's level of wealth, W, follows the process

$$dW \;=\; \sum_{i=1}^{n} \omega_i (\mu_i - r) W \, dt \;+\; (rW - C_t) \, dt \;+\; \sum_{i=1}^{n} \omega_i W \sigma_i \, dz_i \tag{3}$$

where ω_i is the proportion of wealth invested in risky asset i. The habit level, x_s, is assumed to follow the process

$$dx = f\left(\overline{C}_t, x_t\right) dt \tag{4}$$

where $\overline{C}_t$ is the date t consumption that determines the individual's habit.

a. Let $J(W, x, t)$ be the individual's derived utility-of-wealth function. Write down the continuous-time Bellman equation that $J(W, x, t)$ satisfies.

b. Derive the first-order conditions with respect to the portfolio weights, ω_i. Does the optimal portfolio proportion of risky asset i to risky asset j, ω_i/ω_j, depend on the individual's preferences? Why or why not?

c. Assume that the consumption, $\overline{C}_t$, in equation (4) is such that the individual's preferences display an internal habit, similar to the Constantinides model (Constantinides 1990). Derive the first-order condition with respect to the individual's date t optimal consumption, C_t.

d. Assume that the consumption, $\overline{C}_t$, in equation (4) is such that the individual's preferences display an external habit, similar to the Campbell-Cochrane model (Campbell and Cochrane 1999). Derive the first-order condition with respect to the individual's date t optimal consumption, C_t.