Strategic Asset Allocation

John Y. Campbell and Luis M. Viceira

November 2000

CRACKED TRADING SOFTWARE

70+ DVD's FOR SALE & EXCHANGE

www.traders-software.com

www.forex-warez.com

www.trading-software-collection.com

www.tradestation-download-free.com

Contacts

andreybbrv@gmail.com
andreybbrv@yandex.ru
Skype: andreybbrv

Preface

This book has its origins in John Campbell's graduate class in asset pricing at Harvard University in 1995–96. The course emphasized a simplified approach to the difficult problem of intertemporal asset pricing, in which nonlinear equations are approximated by loglinear equations that capture much of the economics of the problem. Luis Viceira, the teaching fellow for the course, proposed that a similar approach could be used to study portfolio choice. This became the basis of Viceira's Harvard PhD thesis, on labor income risk and portfolio choice, and of a series of joint papers studying various types of asset risk: the risk of a changing equity premium (Campbell and Viceira 1999), the risks of changing real interest rates and inflation (Campbell and Viceira 2000), and the risk of changing volatility (Chacko and Viceira 1999). At the same time, a number of other financial economists realized that portfolio choice theory, long a rather quiet backwater of finance, was again an exciting frontier. Papers by Kim and Omberg (1996), Brennan, Schwartz, and Lagnado (1997), Brennan (1998), Barberis (1999), Brandt (1999) and others have radically altered our understanding of this important subject.

In 1999, Campbell delivered three Clarendon Lectures at Oxford summarizing a large part of this literature. His first lecture covered the material in Chapters 1 through 3 of this book; his second lecture presented the material in Chapter 4; and his last lecture discussed material in Chapters 6 and 7. This book further expands the scope of the discussion in an attempt to survey all the major themes of the portfolio choice literature in the years leading up to 2000.

There has always been a tension in economics between the attempt to describe the optimal choices of fully rational individuals ("positive economics") and the desire to use our models to improve people's imperfect choices ("normative economics"). The desire to improve the world with economics was well expressed by Keynes (1930): "If economists could manage to get themselves thought of as humble, competent people, on a level with dentists, that would be splendid!" For much of the 20th Century, economists concentrated on improving economic outcomes through government economic policy; Keynes may have imagined the economist as orthodontist, intervening with the painful but effective tools of monetary and fiscal policy. Today, dentists spend much of their time giving advice on oral hygiene; similarly, economists can try to provide useful advice to improve the myriad economic decisions that private individuals are asked to make. This book represents an attempt at normative economics of this sort.

Of course, optimal portfolio decisions depend on the details of the environment: the financial assets that are available, their expected returns and risks, and the preferences and circumstances of investors. These details become particularly important for long-term investors, who are the subject of this book. Such investors must concern themselves not only with expected returns and risks today, but with the way in which expected returns and risks may change over time. They must also consider their income today and their income prospects for the future. Accordingly this book emphasizes the statistical analysis of asset returns and of income.

Academic economists are not the only, or even the leading source of financial advice for long-term investors. A sizeable financial planning industry has arisen to help people save for retirement. This industry is highly sophisticated in some respects (for example in tax planning), but tends to rely on rules of thumb to guide the tradeoff between risk and return. Conservative investors, for example, are advised to hold fewer equities and more bonds than aggressive investors; younger investors are told that it is appropriate for them to take greater equity risk than older investors. An important purpose of this book is to evaluate such rules of thumb and place them on a firm scientific foundation.

Chapter 1

Introduction

One of the most important decisions many people face is the choice of a portfolio of assets for retirement savings. These assets may be held as a supplement to defined-benefit public or private pension plans; or they may be accumulated in a defined-contribution pension plan, as the major source of retirement income. In either case, a dizzying array of assets is available.

Consider for example the increasing set of choices offered by TIAA-CREF, the principal pension organization for university employees in the United States. Until 1988, the two available choices were TIAA, a traditional nominal annuity, and CREF, an actively managed equity fund. Funds could readily be moved from CREF to TIAA, but the reverse transfer was difficult and could only be accomplished gradually. In 1988, it became possible to move funds between two CREF accounts, a money market fund and an equity fund. Since then, other choices have been added: a bond fund and a socially responsible stock fund in 1990, a global equity fund in 1992, equity index and growth funds in 1994, a real estate fund in 1995, and an inflation-indexed bond fund in 1997. Retirement savings can easily be moved among these funds, each of which represents a broad class of assets with a different profile of returns.

Institutional investors also face complex decisions. Some institutions invest on behalf of their clients, but others, such as foundations and university endowments, are more similar to individuals in that they seek to finance a long-term stream of discretionary spending. The investment options for these institutions have also expanded enormously since the days when a portfolio of government bonds was the norm.

What does financial economics have to say about these investment decisions? Modern finance theory is often thought to have started with the mean-variance analysis of Markowitz (1952); this makes portfolio choice theory the original subject of modern finance. Markowitz showed how investors should pick assets if they care only about the mean and variance—or equivalently the mean and standard deviation—of portfolio returns over a single period.

The results of his analysis are shown in the standard mean-standard deviation diagram, Figure 1.1. (A much more careful mathematical explanation can be found in the next chapter.) For simplicity the figure considers three assets, stocks, bonds, and cash (not literally currency, but a short-term money market fund). The vertical axis shows expected return, and the horizontal axis shows risk as measured by standard deviation. Stocks are shown as offering a high mean return and a high standard deviation, bonds a lower mean and lower standard deviation. Cash has a lower mean return again, but is riskless over one period, so it is plotted on the vertical zero-risk axis. (In the presence of inflation risk, nominal money market investments are not literally riskless in real terms, but this short-term inflation risk is small enough that it is conventional to ignore it. We follow this convention here and return to the issue in the next chapter.)

The curved line in Figure 1.1 shows the set of means and standard deviations that can be achieved by combining stocks and bonds in a risky portfolio. When cash is added to a portfolio of risky assets, the set of means and standard deviations that can be achieved is a straight line on the diagram connecting cash to the risky portfolio. An investor who cares only about the mean and standard deviation of his portfolio will choose a point on the straight line illustrated in the figure, that is tangent to the curved line. This straight line, the mean-variance efficient frontier, offers the highest mean return for any given standard deviation. The point where the straight line touches the curved line is a "tangency portfolio" of risky assets, marked in the figure as "Best Mix of Stocks and Bonds".

The striking conclusion of this analysis is that all investors who care only about mean and standard deviation will hold the same portfolio of risky assets, the unique best mix of stocks and bonds. Conservative investors will combine this portfolio with cash to achieve a point on the mean-variance efficient frontier that is low down and to the left; moderate investors will reduce their cash holdings, moving up and to the right; aggressive investors may even borrow to leverage their holdings of the tangency portfolio, reaching a point on the straight line that is even riskier than the tangency portfolio. But none of these investors should alter the relative proportions of risky assets in the tangency portfolio. This result is the mutual fund theorem of Tobin (1958).

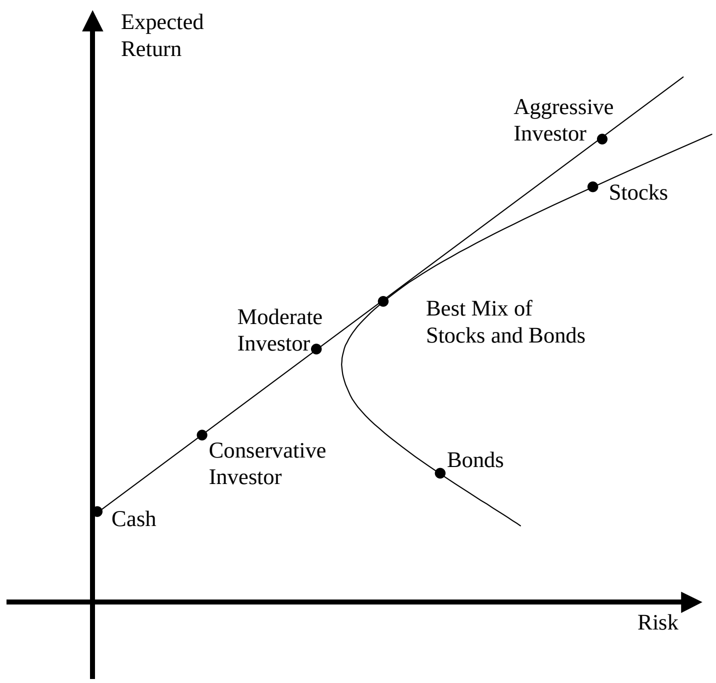

Figure 1.1: Mean-standard deviation diagram

Financial planners have traditionally resisted the simple investment advice embodied in Figure 1.1. This resistance may to some extent be self-serving; as Peter Bernstein points out in his 1992 book *Capital Ideas*, many financial planners and advisers justify their fees by emphasizing the need for each investor to build a portfolio reflecting his or her unique personal situation. Bernstein calls this the "interior decorator fallacy", the view that portfolios should reflect personal characteristics in the same way that interior decor reflects personal taste.[1]

Financial planners' advice does however follow some systematic patterns, and these patterns should be treated with respect. We shall argue in this book that the traditional academic analysis of portfolio choice needs to be modified to handle long investment horizons and labor income; the necessary modifications partially justify, but also qualify, several of the patterns that we see in conventional financial planning advice.

One strong pattern is the tendency for financial planners to encourage young investors, with a long investment horizon, to take more risk than older investors. The single-period mean-variance analysis illustrated in Figure 1.1 assumes a short investment horizon. In this book we shall explore the conditions under which a long investment horizon indeed justifies greater risk-taking.

A second pattern in financial planning advice is that conservative investors are typically encouraged to hold more bonds, relative to stocks, than aggressive investors, contrary to the constant bond-stock ratio illustrated in Figure 1.1. Canner, Mankiw, and Weil (1997) call this the *asset allocation puzzle*. Figure 1.2, which reproduces Table 1 from Canner, Mankiw, and Weil's article, illustrates the asset allocation puzzle. The table summarizes model portfolios recommended by four different investment advisers in the early 1990's: Fidelity, Merrill Lynch, the financial journalist Jane Bryant Quinn, and the *New York Times*. While the portfolios differ in their details, in every case the recommended ratio of bonds to stocks is higher for moderate investors than for aggressive investors, and higher again for conservative investors.

[1] An amusing recent example is the PaineWebber advertisement that ran in the *New Yorker* in 1998: "If our clients were all the same, their *portfolios* would be too. *They say* the research has been sifted. The numbers have been crunched. The analysts have spoken: Behold! The ideal portfolio. *We say* building a portfolio is not 'one size fits all.' It begins with knowing you—how you feel about money, how much risk you can tolerate, your hopes for your family, and for your future. By starting with the human element, our Financial Advisors can do something a black box can't do—take the benefits of what PaineWebber has to offer and create an investment plan unique to you."

CANNER ET AL.: AN ASSET ALLOCATION PUZZLE

TABLE 1—ASSET ALLOCATIONS RECOMMENDED BY FINANCIAL ADVISORS

	Percent of portfolio			
Advisor and investor type	Cash	Bonds	Stocks	Ratio of bonds to stocks
A. Fidelity[a]				
Conservative	50	30	20	1.50
Moderate	20	40	40	1.00
Aggressive	5	30	65	0.46
B. Merrill Lynch[b]				
Conservative	20	35	45	0.78
Moderate	5	40	55	0.73
Aggressive	5	20	75	0.27
C. Jane Bryant Quinn[c]				
Conservative	50	30	20	1.50
Moderate	10	40	50	0.80
Aggressive	0	0	100	0.00
D. *The New York Times*[d]				
Conservative	20	40	40	1.00
Moderate	10	30	60	0.50
Aggressive	0	20	80	0.25

Sources:
[a] Mark, 1993.
[b] Underwood and Brown, 1993.
[c] Quinn, 1991.
[d] Rowland, 1994.

Figure 1.2: The Asset Allocation Puzzle

One possible explanation for this pattern of advice is that aggressive investors are unable to borrow at the riskless interest rate, and thus cannot reach the upper right portion of the straight line in Figure 1.1. In this situation, aggressive investors should move along the curved line, increasing their allocation to stocks and reducing their allocation to bonds. The difficulty with this explanation is that it only applies once the constraint on borrowing starts to bind on investors, that is, once cash holdings have been reduced to zero; but the bond-stock ratio in Canner, Mankiw, and Weil's Table 1 varies even when cash holdings are positive. Elton and Gruber (2000) respond to this difficulty by arguing that these cash holdings are a special liquidity reserve determined outside the mean-variance analysis. It is quite plausible that the 5% cash holdings suggested by Fidelity and Merrill Lynch are a liquidity reserve, but the other cash holdings in the table appear too large to be explained in this manner.

This book argues that it is possible to make sense of both the asset allocation puzzle, and the tendency of financial planners to recommend riskier portfolios to young investors. The key is to recognize that optimal portfolios for long-term investors need not be the same as for short-term investors. Long-term investors, who value wealth not for its own sake but for the standard of living that it can support, may judge risks very differently from short-term investors. Cash, for example, is risky in the long term even though it is safe in the short term, because cash holdings must be reinvested in the future at unknown real interest rates. Inflation-indexed bonds, on the other hand, provide a known stream of long-term real payments even though their capital value is uncertain in the short term. There is considerable evidence that stocks, too, can support a stable standard of living more successfully than their short-term price variability would indicate. For these reasons a long-term investor may be willing to hold more stocks and bonds, and less cash, than a short-term investor would do; and a conservative long-term investor may hold a portfolio that is dominated by bonds rather than cash.

Labor income is also important for long-term investors. One can think of working investors as implicitly holding an asset, human capital, whose dividends equal labor income. This asset is nontradable, so investors cannot sell it; but they can adjust their financial asset holdings to take account of their implicit holdings of human capital. For most investors, human capital is sufficiently stable in value to tilt financial portfolios towards greater holdings of risky assets.

The rest of the book is organized as follows. Chapter 2 first reviews the traditional mean-variance analysis, showing how it can be founded on utility theory. The chapter argues that the benchmark model of utility should assume that relative risk aversion is independent of wealth. With this assumption, there are well-known conditions under which long-term investors should invest myopically, choosing the same portfolios as short-term investors. Chapter 2 explains these conditions, due originally to Merton (1969) and Samuelson (1969): Myopic portfolio choice is optimal if investors have no labor income and investment opportunities are constant over time. If investors have relative risk aversion equal to one, then myopic portfolio choice is optimal even if investment opportunities are time-varying. Although these conditions are simple, they are widely misunderstood and the chapter makes an effort to address fallacies that commonly arise in popular discussion.

Legitimate arguments for horizon effects on portfolio choice depend on violations of the Merton-Samuelson conditions. Chapters 3 through 7 explore such violations. Chapter 3 argues that investment opportunities are not constant because real interest rates move over time. Even if expected excess returns on risky assets over safe assets are constant, time-variation in real interest rates is enough to generate large differences between optimal portfolios for long-term and short-term investors. The chapter shows that conservative long-term investors should hold portfolios that consist largely of long-term bonds. These bonds should be inflation-indexed if possible; however nominal bonds may be adequate substitutes for inflation-indexed bonds if inflation risk is modest, as it has been in the United States since the early 1980's.

The assumption of constant risk premia in Chapter 3 implies that optimal portfolios are constant over time for both short-term and long-term investors. Chapter 4 allows for time-variation in the expected excess returns on stocks and bonds, which generates time-variation in optimal portfolios. Both short-term and long-term investors should seek to "time the markets", holding more risky assets at times when the rewards for doing so are high. But in addition, long-term investors with relative risk aversion greater than one should increase their average holdings of risky assets whose returns are negatively correlated with the rewards for riskbearing; for example, they should increase their average allocation to stocks because the stock market appears to mean-revert, doing relatively poorly after price increases and relatively well after price declines. These findings are an empirical development of Merton's (1973) theoretical concept of intertemporal hedging by long-term investors.

Chapter 5 seeks to relate the results of Chapters 3 and 4 more closely to the extensive theoretical literature set in continuous time. Explicit solutions for optimal portfolios are provided in Chapters 3 and 4 by the use of loglinear approximations to discrete-time Euler equations and budget constraints. Chapter 5 clarifies the conditions under which such approximate solutions hold exactly, and shows how equivalent approximations can be used in continuous time. This chapter also explores optimal portfolio choice in the presence of time-varying stock market risk. Chapter 5 is the most technically demanding chapter in the book, and less mathematical readers can skip it without loss of continuity.

Chapters 6 and 7 introduce labor income into the long-term portfolio choice problem. Chapter 6 discusses labor income in a stylized two-period model and a fairly abstract infinite-horizon setting, while Chapter 7 embeds labor income in a life-cycle model and asks how investors should adjust their portfolios as they age. This chapter also reviews the existing empirical evidence on how investors actually do invest over the life cycle. The current draft of this book omits Chapters 6 and 7.

Chapter 2

Myopic Portfolio Choice

In this chapter we review the theory of portfolio choice for short-term investors, and explain those special cases in which long-term investors should make the same choices as short-term investors. In these special cases the investment horizon is irrelevant; portfolio choice is said to be myopic, because investors ignore what will happen beyond the immediate next period. Throughout the chapter we assume that investors have financial wealth but no labor income.

Section 2.1 describes optimal portfolio choice for short-term investors. We begin in section 2.1.1 with the classic mean-variance analysis, assuming that investors care directly about the mean and variance of portfolio returns over one period. Then in section 2.1.2 we derive similar results assuming that investors have a utility function defined over wealth at the end of one period. We discuss alternative assumptions that can be made about utility, arguing that there are good reasons to prefer scale-independent utility functions in which relative risk aversion does not depend on wealth. The simplest scale-independent utility function is power utility, and we show how to derive portfolio results analogous to those of the mean-variance analysis, assuming power utility and lognormally distributed returns.

In section 2.2 we derive conditions under which the same portfolio choice is optimal for long-term investors. We first assume in section 2.2.1 that investors have power utility defined over wealth many periods ahead, and show that if investors can rebalance their portfolios each period, they should invest myopically if asset returns are independent and identically distributed (IID) over time, or if utility takes the log form. Log utility is the special case of power utility in which both the coefficient of relative risk aversion and the elasticity of intertemporal substitution in consumption equal one.

The conditions for myopic portfolio choice are quite simple, and can be derived without the use of advanced mathematics. Nonetheless these conditions are widely misunderstood, and one often sees specious arguments that there should be horizon effects even when the conditions hold. In section 2.2.2 we try to expose the fallacies in these arguments.

In section 2.2.3 we consider investors who have power utility defined over consumption, and show that portfolio choice will be myopic under the same conditions as before. Finally, in section 2.2.4 we introduce a generalization of power utility, Epstein-Zin utility, which allows us to distinguish between the coefficient of risk aversion and the elasticity of intertemporal substitution in consumption. Power utility links these concepts tightly together, making one the reciprocal of the other, but they are different concepts that play quite different roles in the analysis. We show that portfolio choice is myopic if relative risk aversion equals one, regardless of the value of the elasticity of intertemporal substitution in consumption. Epstein-Zin utility will be used extensively in the rest of the book.

Throughout the chapter there is a strong emphasis on the difference between simple returns and log returns, and on the adjustments that are needed to translate from one type of return to the other. Elementary treatments of portfolio choice often gloss over this difference, but it is central to the theory of portfolio choice for long-term investors.

2.1 Short-term portfolio choice

2.1.1 Mean-variance analysis

Choosing the weight on a single risky asset

Consider the following classic portfolio choice problem. Two assets are available to an investor at time t. One is riskless, with simple return $R_{f,t+1}$ from time t to time $t+1$, and one is risky. The risky asset has simple return R_{t+1} from time t to time $t+1$, with conditional mean $\mathrm{E}_t R_{t+1}$ and conditional variance σ_t^2. Note the timing convention that returns are given time subscripts for the date at which they are realized; the riskfree interest rate is realized at $t+1$ but is known one period in advance at time t. The conditional mean and conditional variance are the mean and variance conditional on the investor's information at time t, thus they are given t subscripts.

The investor puts a share α_t of his portfolio into the risky asset. Then the portfolio return is

$$R_{p,t+1} = \alpha_t R_{t+1} + (1-\alpha_t)R_{f,t+1} = R_{f,t+1} + \alpha_t(R_{t+1} - R_{f,t+1}). \tag{2.1}$$

The mean portfolio return is $\mathrm{E}_t R_{p,t+1} = R_{f,t+1} + \alpha_t(\mathrm{E}_t R_{t+1} - R_{f,t+1})$, while the variance of the portfolio return is $\sigma_{pt}^2 = \alpha_t^2 \sigma_t^2$.

The investor prefers a high mean and a low variance of portfolio returns. We assume that the investor trades off mean and variance in a linear fashion. That is, the investor maximizes a linear combination of mean and variance, with a positive weight on mean and a negative weight on variance:

$$\max_{\alpha_t}(\mathrm{E}_t R_{p,t+1} - \frac{k}{2}\sigma_{pt}^2). \tag{2.2}$$

Substituting in the mean and variance of portfolio returns, and subtracting $R_{f,t+1}$ (which does not change the maximization problem), this can be rewritten as

$$\max_{\alpha_t} \alpha_t(\mathrm{E}_t R_{t+1} - R_{f,t+1}) - \frac{k}{2}\alpha_t^2\sigma_t^2. \tag{2.3}$$

The solution to this maximization problem is

$$\alpha_t = \frac{\mathrm{E}_t R_{t+1} - R_{f,t+1}}{k\sigma_t^2}. \tag{2.4}$$

The portfolio share in the risky asset should equal the expected excess return, or risk premium, divided by conditional variance times the coefficient k that represents aversion to variance. We will see similar expressions frequently in this book.

A useful concept in portfolio analysis is the Sharpe ratio S_t, defined as the ratio of mean excess return to standard deviation:

$$S_t = \frac{\mathrm{E}_t R_{t+1} - R_{f,t+1}}{\sigma_t}. \tag{2.5}$$

The portfolio solution can be rewritten as

$$\alpha_t = \frac{S_t}{k\sigma_t}. \tag{2.6}$$

The mean excess return on the portfolio is S_t^2/k and the variance of the portfolio is S_t^2/k^2, so the ratio of mean to variance is $1/k$. The standard deviation of the portfolio is S_t/k, and so the Sharpe ratio of the portfolio is S_t. In this simple model, all portfolios have the same Sharpe ratio because they all contain the same risky asset in greater or smaller amount.

Mean-variance analysis with many risky assets

These results extend straightforwardly to the case where there are many risky assets. We define the portfolio return in the same manner as before, except that we use lowercase boldface letters to denote vectors and uppercase boldface letters to denote matrices. Thus $\mathbf{R}_{t+1}$ is now a vector of risky returns with N elements. It has a mean vector $\mathrm{E}_t\mathbf{R}_{t+1}$ and a variance-covariance matrix $\mathbf{\Sigma}_t$. Also, $\boldsymbol{\alpha}_t$ is now a vector of allocations to the risky assets. The maximization problem (2.3) now becomes

$$\max_{\boldsymbol{\alpha}_t} \boldsymbol{\alpha}_t'(\mathrm{E}_t\mathbf{R}_{t+1} - R_{f,t+1}\boldsymbol{\iota}) - \frac{k}{2}\boldsymbol{\alpha}_t'\mathbf{\Sigma}_t\boldsymbol{\alpha}_t. \tag{2.7}$$

Here $\boldsymbol{\iota}$ is a vector of ones, and $(\mathrm{E}_t\mathbf{R}_{t+1} - R_{f,t+1}\boldsymbol{\iota})$ is the vector of excess returns on the N risky assets over the riskless interest rate. The variance of the portfolio return is $\boldsymbol{\alpha}_t'\mathbf{\Sigma}_t\boldsymbol{\alpha}_t$.

The solution to this maximization problem is

$$\boldsymbol{\alpha}_t = \frac{1}{k}\mathbf{\Sigma}_t^{-1}(\mathrm{E}_t\mathbf{R}_{t+1} - R_{f,t+1}\boldsymbol{\iota}). \tag{2.8}$$

This is a straightforward generalization of the solution with a single risky asset. The single excess return is replaced by a vector of excess returns, and the reciprocal of variance is replaced by $\mathbf{\Sigma}_t^{-1}$, the inverse of the variance-covariance matrix of returns.

The investor's preferences enter the solution (2.8) only through the scalar term $1/k$. Thus investors differ only in the overall scale of their risky asset position, not in the composition of that position. Conservative investors with a high k hold more of the riskless asset and less of all risky assets, but they do not change the relative proportions of their risky assets which are determined by the vector $\mathbf{\Sigma}_t^{-1}(\mathrm{E}_t\mathbf{R}_{t+1} - R_{f,t+1}\boldsymbol{\iota})$. This is the mutual fund theorem of Tobin (1958), as illustrated in Figure 1.1.

The results also extend straightforwardly to the case where there is no completely riskless asset. We can still define a benchmark asset with return $R_{0,t+1}$, and define excess returns relative to this benchmark return. The variance of the portfolio return is $\mathrm{Var}_t(R_{0,t+1}) + \boldsymbol{\alpha}_t'\mathbf{\Sigma}_t\boldsymbol{\alpha}_t + 2\boldsymbol{\alpha}_t'\boldsymbol{\sigma}_{0t}$, where $\mathbf{\Sigma}_t$ is now defined to be the conditional variance-covariance matrix of excess returns over the benchmark asset, and $\boldsymbol{\sigma}_{0t}$ is a vector containing the covariances of the excess returns on the other assets with the benchmark return. (We use a boldface lowercase rather than a boldface uppercase sigma for $\boldsymbol{\sigma}_{0t}$ because this is a vector rather than a matrix. Throughout the book we will use a boldface lowercase sigma, with a suitable subscript, to denote a vector of covariances between excess returns and the subscripted variable. Abusing this notation slightly, we will also write $\boldsymbol{\sigma}_t^2$ for the vector of excess-return variances.)

With no riskless asset, the solution becomes

$$\boldsymbol{\alpha}_t = \frac{1}{k}\boldsymbol{\Sigma}_t^{-1}(\mathrm{E}_t\mathbf{R}_{t+1} - R_{0,t+1}\boldsymbol{\iota}) - \boldsymbol{\Sigma}_t^{-1}\boldsymbol{\sigma}_{0t}. \tag{2.9}$$

This has almost the same form as before, except that the relation between portfolio weights and average excess returns is now linear rather than proportional. The intercept is the minimum-variance portfolio of all assets, $-\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\sigma}_{0t}$, which does not place 100% weight in the benchmark asset if the benchmark asset is risky. It is no longer true that all investors hold risky assets in the same proportions; instead they hold some combination of two risky mutual funds, whose proportions are given by the two terms on the right-hand side of (2.9). If the benchmark asset has low risk, however, as is the case empirically for Treasury bills and other short-term debt instruments, then there is little difference between the solution (2.9) and the riskless-asset solution (2.8).

2.1.2 Specifying utility of wealth

Basics of utility theory

So far we have assumed that investors care directly about the mean and the variance of portfolio returns. Similar results are available if we assume instead that investors have utility defined over wealth at the end of the period. In this case we redefine the maximization problem as

$$\max \mathrm{E}_t U(W_{t+1}) \tag{2.10}$$

subject to

$$W_{t+1} = (1 + R_{p,t+1})W_t. \tag{2.11}$$

Here $U(W_{t+1})$ is a standard concave utility function, as illustrated in Figure 2.1. The curvature of the utility function implies that the investor is averse to risk. Consider for example an investor with initial wealth W_t who is offered a risky gamble that will either add or subtract an amount G to wealth, with equal probabilities of the two outcomes. If the investor turns down the gamble, wealth is certain and utility is $U(W_t)$. If the investor accepts the gamble, there is a one-half chance that wealth will go up to $W_t + G$ and a one-half chance that it will fall to $W_t - G$. Expected utility is

$(1/2)U(W_t+G)+(1/2)U(W_t-G)$, which is less than $U(W_t)$ because of the curvature of the utility function. Thus the investor turns down the gamble; it offers only risk without any accompanying reward, and is unattractive to a risk-averse investor.

The degree of curvature of the utility function determines the intensity of the investor's risk aversion. Curvature can be measured by the second derivative of the utility function with respect to wealth, scaled by the first derivative to eliminate any dependence of the measure of curvature on the arbitrary units in which utility is measured. The *coefficient of absolute risk aversion* is then defined as

$$A(W) = -\frac{U''(W)}{U'(W)}, \tag{2.12}$$

and the *coefficient of relative risk aversion* is defined as

$$R(W) = WA(W) = -\frac{WU''(W)}{U'(W)}. \tag{2.13}$$

The reciprocals of these measures are called *absolute and relative risk tolerance.*

Classic results of Pratt (1964) say that for small gambles, the coefficient of absolute risk aversion determines the absolute dollar amount that an investor is willing to pay to avoid a gamble of a given absolute size. It is commonly thought that absolute risk aversion should decrease, or at least should not increase, with wealth. Introspection suggests that a billionaire will be relatively unconcerned with a risk that might worry a poor person, and will pay less to avoid such a risk.

The coefficient of relative risk aversion determines the fraction of wealth that an investor will pay to avoid a gamble of a given size relative to wealth. A plausible benchmark model makes relative risk aversion independent of wealth. In this case people at all levels of wealth make the same decisions, when both risks and the costs of avoiding them are expressed as fractions of wealth.

The long-run behavior of the economy suggests that relative risk aversion cannot depend strongly on wealth. Per capita consumption and wealth have increased greatly over the past two centuries. Since financial risks are multiplicative, this means that the absolute scale of financial risks has also increased while the relative scale is unchanged. Interest rates and risk premia do not show any evidence of long-term trends in response to this long-term growth; this implies that investors are willing to pay almost

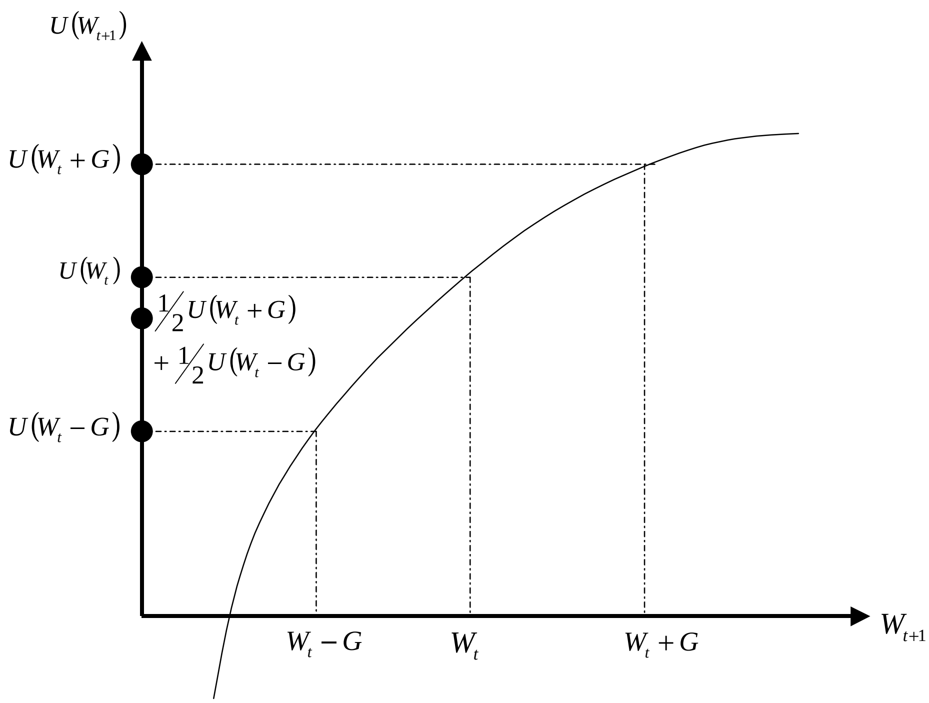

Figure 2.1: Concave utility of wealth

the same relative costs to avoid given relative risks as they did when they were much poorer, which is only possible if relative risk aversion is almost independent of wealth.

The form of the utility function

Tractable models of portfolio choice require assumptions about the form of the utility function, and possibly distributional assumptions about asset returns. Three alternative sets of assumptions produce simple results that are consistent with those of mean-variance analysis:

1. Investors have quadratic utility defined over wealth. In this case $U(W_{t+1}) = a + bW_{t+1}$. Under this assumption maximizing expected utility, as in (2.10), is equivalent to maximizing a linear combination of mean and variance, as in (2.2). No distributional assumptions are needed on asset returns. Quadratic utility implies that absolute risk aversion and relative risk aversion are increasing in wealth.

2. Investors have exponential utility, $U(W_{t+1}) = -\exp(-\theta W_{t+1})$, and asset returns are normally distributed. Exponential utility implies that absolute risk aversion is a constant θ, while relative risk aversion increases in wealth.

3. Investors have power utility, $U(W_{t+1}) = (W_{t+1}^{1-\gamma} - 1)/(1-\gamma)$, and asset returns are lognormally distributed. Power utility implies that absolute risk aversion is declining in wealth, while relative risk aversion is a constant γ. The limit as γ approaches one is log utility: $U(W_{t+1}) = \log(W_{t+1})$.

We have already argued that absolute risk aversion should decline, or at the very least should not increase, with wealth. This rules out the assumption of quadratic utility, and favors power utility over exponential utility. The power-utility property of constant relative risk aversion is inherently attractive, and is required to explain the stability of financial variables in the face of secular economic growth.

The choice between exponential and power utility also implies distributional assumptions on returns. Exponential utility produces simple results if asset returns are normally distributed, while power utility produces simple results if asset returns are lognormal (that is, if their logs are normal).

The assumption of normal returns is appealing for some purposes, but it is inappropriate for the study of long-term portfolio choice because it cannot hold at more than one time horizon. If returns are normally distributed at a monthly frequency, then two-month returns are not normal because they are the product of two successive normal returns and sums of normals, not products of normals, are themselves normal. The assumption of lognormal returns, on the other hand, can hold at every time horizon since products of lognormal random variables are themselves lognormal.

The assumption of lognormal returns runs into another difficulty, however. It does not carry over straightforwardly from individual assets to portfolios. A portfolio is a linear combination of individual assets; if each asset return is lognormal, the portfolio return is a weighted average of lognormals which is not itself lognormal. This difficulty can be avoided by considering short time intervals. As the time interval shrinks, the non-lognormality of the portfolio return diminishes, and it disappears altogether in the limit of continuous time. In this and the next few chapters we use a discrete-time approximation to the relation between the log return on a portfolio and the log returns on individual assets. The approximation becomes more accurate as the time interval shrinks. In Chapter 5 we develop explicit models set in continuous time.

2.1.3 A lognormal model with power utility

We now develop portfolio choice results under the assumption that investors have power utility and that asset returns are lognormal. We repeatedly apply a key result about the expectation of a lognormal random variable X:

$$\log \mathrm{E}_t X_{t+1} = \mathrm{E}_t \log X_{t+1} + \frac{1}{2}\mathrm{Var}_t \log X_{t+1} = \mathrm{E}_t x_{t+1} + \frac{1}{2}\sigma_{xt}^2. \tag{2.14}$$

Here and throughout the book, the notation log refers to the natural logarithm, and lower-case letters are used to denote the logs of the corresponding upper-case letters. Equation (2.14) can be understood intuitively by reference to Figure 2.1. The log is a concave function like the utility function illustrated in Figure 2.1. Thus the mean of the log of a random variable X is smaller than the log of the mean, and the difference is increasing in the variability of X. The equation quantifies this difference for the special case in which $\log X$ is normally distributed.

Under the assumption of power utility, equation (2.10) can be written as

$$\max \mathrm{E}_t W_{t+1}^{1-\gamma}/(1-\gamma). \tag{2.15}$$

Maximizing this expectation is equivalent to maximizing the log of the expectation, and the scale factor $1/(1-\gamma)$ can be omitted since it does not affect the solution. Under the assumption that next-period wealth is lognormal, we can apply equation (2.14) to rewrite as

$$\max \log \mathrm{E}_t W_{t+1}^{1-\gamma} = (1-\gamma)\mathrm{E}_t w_{t+1} + \frac{1}{2}(1-\gamma)^2\sigma_{wt}^2. \tag{2.16}$$

The budget constraint (2.11) can be rewritten in log form as

$$w_{t+1} = r_{p,t+1} + w_t, \tag{2.17}$$

where $r_{p,t+1} = \log(1 + R_{p,t+1})$ is the log return on the portfolio, the natural logarithm of the gross simple return, also known as the continuously compounded portfolio return. Dividing (2.16) by $(1-\gamma)$ and using (2.17), we restate the problem as

$$\max \mathrm{E}_t r_{p,t+1} + \frac{1}{2}(1-\gamma)\sigma_{pt}^2, \tag{2.18}$$

where σ_{pt}^2 is the conditional variance of the log portfolio return.

To understand this equation, it is helpful to note that

$$E_t r_{p,t+1} + \sigma_{pt}^2/2 = \log E_t(1 + R_{p,t+1}) \tag{2.19}$$

because the portfolio return is lognormal. Thus (2.18) can be rewritten as

$$\max \log \mathrm{E}_t(1 + R_{p,t+1}) - \frac{1}{2}\gamma\sigma_{pt}^2. \tag{2.20}$$

Just as in the mean-variance analysis, the investor trades off mean against variance in the portfolio return. The relevant mean return is the mean simple return, or arithmetic mean return, and the investor trades the log of this mean linearly against the variance of the log return. The coefficient of relative risk aversion, γ, plays the same role here as the parameter k played in the mean-variance analysis.

Equation (2.18) shows that the case $\gamma = 1$ plays a special role in the analysis. When $\gamma = 1$, the investor has log utility and chooses the portfolio with the highest available log return (sometimes known as the "growth-optimal" portfolio). When $\gamma > 1$, the investor seeks a safer portfolio by penalizing the variance of log returns; when $\gamma < 1$, the investor actually seeks a riskier portfolio because a higher variance, with the same mean log return, corresponds to a higher mean simple return. The case $\gamma = 1$ is the boundary where these two opposing considerations exactly cancel one another out. This case plays an important role throughout the book.

Approximation of the portfolio return

To proceed further, we need to relate the log portfolio return to the log returns on underlying assets. Consider first the simple case where there are two assets, one risky and one riskless. Then from (2.1) the simple return on the portfolio is a linear combination of the simple returns on the risky and riskless assets. The log return on the portfolio is the log of this linear combination, which is not the same as a linear combination of logs.

Over short time intervals, however, we can use a Taylor approximation of the nonlinear function relating log individual-asset returns to log portfolio returns. Full details are given in the Appendix; the resulting expression is

$$r_{p,t+1} - r_{f,t+1} = \alpha_t(r_{t+1} - r_{f,t+1}) + \frac{1}{2}\alpha_t(1-\alpha_t)\sigma_t^2. \tag{2.21}$$

The difference between the log portfolio return and a linear combination of log individual-asset returns is given by $\alpha_t(1-\alpha_t)\sigma_t^2/2$. The difference disappears if the portfolio weight in the risky asset is zero (for then the log portfolio return is just the log riskless return), or if the weight in the risky asset is one (for then the log portfolio return is just the log risky return). When $0 < \alpha_t < 1$, the portfolio is a weighted average of the individual assets and the term $\alpha_t(1-\alpha_t)\sigma_t^2/2$ is positive. To understand this, recall that the log of an average is greater than an average of logs as illustrated in Figure 2.1.

The approximation in (2.21) can be justified rigorously by considering shorter and shorter time intervals. As the time interval shrinks, the higher-order terms that are neglected in (2.21) become negligible relative to those that are included. In the limit of continuous time with continuous paths (diffusions) for asset prices, (2.21) is exact and can be derived using Ito's Lemma. We discuss the continuous-time approach in more detail in Chapter 5.

One important property of the approximate portfolio return is that it rules out the possibility of bankruptcy, even when the investor holds a short position ($\alpha_t < 0$) or a leveraged position in the risky asset financed by borrowing ($\alpha_t > 1$). The log portfolio return is always finite, no matter what the returns on the underlying assets, and thus it is never possible to exhaust wealth completely. Continuous-time diffusion models also have this property; in such models portfolios are rebalanced over such short intervals that losses can always be stemmed by rebalancing before they lead to bankruptcy. In most applications it is reasonable to exclude the possibility of bankruptcy, but this approach may not be suitable when it implies optimal portfolios with extremely high leverage.

The approximation (2.21) generalizes straightforwardly to the case where there are many risky assets, jointly lognormally distributed with conditional variance-covariance matrix of log returns $\mathbf{\Sigma}_t$. We write $\boldsymbol{\sigma}_t^2$ for the vector containing the diagonal elements of $\mathbf{\Sigma}_t$, the variances of asset returns. (Recall our notational conventions that boldface lowercase letters denote vectors, boldface lowercase sigma denotes a vector of covariances, and with a slight abuse of notation $\boldsymbol{\sigma}_t^2$ denotes a vector of variances.) The approximation to the portfolio return becomes

$$r_{p,t+1} - r_{f,t+1} = \boldsymbol{\alpha}_t'(\mathbf{r}_{t+1} - r_{f,t+1}\boldsymbol{\iota}) + \frac{1}{2}\boldsymbol{\alpha}_t'\boldsymbol{\sigma}_t^2 - \frac{1}{2}\boldsymbol{\alpha}_t'\mathbf{\Sigma}_t\boldsymbol{\alpha}_t. \tag{2.22}$$

This approximation holds in exactly the same form if we replace the riskless return $r_{f,t+1}$ with a risky benchmark return $r_{0,t+1}$, except that in this case the vector $\boldsymbol{\sigma}_t^2$ and the matrix $\mathbf{\Sigma}_t$ must contain variances and covariances of *excess* returns on the other risky assets over the benchmark return, rather than variances and covariances of total returns on these assets. (Excess and total returns only have the same variances and covariances when they are measured relative to a riskless return.)

Solution of the model

In a two-asset model, equation (2.21) implies that the mean excess portfolio return is $\mathrm{E}_t r_{p,t+1} - r_{f,t+1} = \alpha_t(\mathrm{E}_t r_{t+1} - r_{f,t+1}) + \frac{1}{2}\alpha_t(1-\alpha_t)\sigma_t^2$, while the variance of the portfolio return is $\alpha_t^2\sigma_t^2$. Substituting into the objective function (2.18), the problem becomes

$$\max \alpha_t(\mathrm{E}_t r_{t+1} - r_{f,t+1}) + \frac{1}{2}\alpha_t(1-\alpha_t)\sigma_t^2 + \frac{1}{2}(1-\gamma)\alpha_t^2\sigma_t^2. \tag{2.23}$$

The solution is

$$\alpha_t = \frac{\mathrm{E}_t r_{t+1} - r_{f,t+1} + \sigma_t^2/2}{\gamma\sigma_t^2}. \tag{2.24}$$

This equation is the equivalent, in a lognormal model with power utility, of the mean-variance solution (2.4). The top line is the expected excess log return on the risky asset, with the addition of one-half the variance to convert from log returns to simple returns that are ultimately of concern to the investor. (The formula for the expectation of lognormal random variables implies that $\mathrm{E}_t r_{t+1} - r_{f,t+1} + \sigma_t^2/2 = \log \mathrm{E}_t(1+R_{t+1})/(1+R_{ft})$.) The bottom line is the coefficient of relative risk aversion times the variance of the risky asset return. Thus, just as in a simple mean-variance model, the optimal portfolio weight is the risk premium divided by risk aversion times variance.

In a model with many risky assets, the solution for the vector of optimal portfolio weights is

$$\boldsymbol{\alpha}_t = \frac{1}{\gamma}\boldsymbol{\Sigma}_t^{-1}(\mathrm{E}_t\mathbf{r}_{t+1} - r_{f,t+1}\boldsymbol{\iota} + \boldsymbol{\sigma}_t^2/2). \tag{2.25}$$

This solution is the equivalent of the multiple-asset mean-variance solution (2.8). Like the mean-variance solution, it has the property that the coefficient of relative risk aversion only affects the overall scale of the risky asset position and not its composition. Thus a version of Tobin's mutual fund theorem holds in the lognormal model with power utility.

If there is no truly riskless asset, and we work instead with a risky benchmark return $r_{0,t+1}$, then the solution becomes

$$\boldsymbol{\alpha}_t = \frac{1}{\gamma}\boldsymbol{\Sigma}_t^{-1}(\mathrm{E}_t\mathbf{r}_{t+1} - r_{0,t+1}\boldsymbol{\iota} + \boldsymbol{\sigma}_t^2/2) + \left(1 - \frac{1}{\gamma}\right)(-\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\sigma}_{0t}), \tag{2.26}$$

where $\boldsymbol{\sigma}_{0t}$ is the vector of covariances of excess log returns with the benchmark log return. Just as in the simple mean-variance analysis (2.9), covariances with the benchmark affect the optimal portfolio weights. The investor favors assets with positive covariances because, for given expected log returns, they increase the expected simple return on the portfolio; but the investor dislikes such assets because they increase the risk of the portfolio. The two effects cancel when $\gamma = 1$; in this case the solution takes exactly the same form whether the benchmark asset is riskless or risky. As γ increases the optimal portfolio approaches the minimum-variance portfolio $-\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\sigma}_{0t}$.

2.2 Myopic long-term portfolio choice

2.2.1 Power utility of wealth

So far we have assumed that the investor has a short investment horizon and cares only about the distribution of wealth at the end of the next period. Alternatively, we can assume that the investor cares about the distribution of wealth K periods from now, so that the utility function is $U(W_{t+K})$ rather than $U(W_{t+1})$. We continue to assume that all wealth is reinvested, so the budget constraint takes the form

$$W_{t+K} = (1+R_{pK,t+K})W_t = (1+R_{p,t+1})(1+R_{p,t+2})...(1+R_{p,t+K})W_t. \tag{2.27}$$

Here the notation $(1 + R_{pK,t+K})$ indicates that the portfolio return is measured over K periods, from t to $t + K$. This K-period return is just the product of K successive 1-period returns. Note that this is a cumulative return; to calculate an annualized return one would take the Kth root if the base period is a year, the $(K/4)$th root if the base period is a quarter, and in general the (K/S)th root if there are S base periods in a year. Taking logs, the cumulative log return over K periods is just a sum of K 1-period returns:

$$r_{pK,t+K} = r_{p,t+1} + ... + r_{p,t+K}. \tag{2.28}$$

The annualized log return can be found by dividing by (K/S) if there are S base periods in a year.

The long-term investor's optimal portfolio depends not only on his objective, but also on what he is allowed to do each period. In particular, it depends on whether the investor is allowed to rebalance his portfolio each period, or must choose an allocation at time t without any possibility of asset sales or purchases between t and the horizon $t + K$.

Myopic portfolio choice without rebalancing

We first assume that rebalancing is not possible between t and $t + K$, so that the long-term investor must evaluate K-period returns in the same manner that the short-term investor evaluates single-period returns. We continue to assume that utility takes the power form and that asset returns are conditionally lognormally distributed. For simplicity we return to the case where there is a single risky asset.

We now make a highly restrictive additional assumption, that all asset returns are independent and identically distributed (IID) over time. This implies that the log riskless rate is a constant r_f and the log K-period riskless return is Kr_f; the mean log return on the risky asset is a constant $\mathrm{E}r$ and the mean log K-period return on the risky asset is $K\mathrm{E}r$; the variance of the log return on the risky asset is a constant σ^2 and the risky asset return is serially uncorrelated, so the variance of the log K-period return on the risky asset is just $K\sigma^2$:

$$\mathrm{Var}_t r_{K,t+K} = \mathrm{Var}_t r_{t+1} + ... + \mathrm{Var}_t r_{t+K} = K\sigma^2. \tag{2.29}$$

Here the first equality follows from the absence of serial correlation in risky returns, and the second follows from the constant variance of the risky return.

With IID returns, then, the mapping from single-period log returns to K-period log returns is straightforward. All means and variances for individual assets are scaled up by the same factor K. If we can apply the same approximation (2.21) relating individual asset returns to the portfolio return, this implies that the previous short-term portfolio solution is still optimal for a long-term investor. Intuitively, the optimal portfolio weight on a risky asset is the mean excess log return, plus one-half the log variance to convert from mean log to mean simple return, divided by risk aversion times variance. If both the mean and the variance are multiplied by K, this solution does not change. The argument can readily be extended to the case with multiple risky assets. Both the short-term investor and the long-term investor perceive the same mean-variance diagram, merely scaled up or down by a factor K; thus they choose the same point on the diagram, that is, they choose the same portfolio.

The weakness in this analysis is the assumption that the approximate budget constraint (2.21) applies to a long holding period. Recall that this budget constraint holds exactly in continuous time, and is an accurate approximation over short discrete time intervals; but the quality of the approximation deteriorates if it is applied over long holding periods. For this reason an exact solution to the long-term portfolio choice problem without rebalancing does involve some horizon effects even if risky returns are IID (see for example Barberis (2000)). These effects are small but not negligible.

Myopic portfolio choice with rebalancing

The assumption that a long-horizon investor cannot rebalance his portfolio is superficially appealing because it makes the long-horizon problem formally analogous to the short-horizon problem. Unfortunately it creates a technical difficulty because it invalidates the use of the loglinear budget constraint (2.21). More seriously, this assumption does not describe reality. Investors with long horizons are free to trade assets at any time, and financial intermediaries exist to rebalance portfolios on behalf of investors who find this task costly to execute. (This is the purpose of so-called lifestyle mutual funds.) There is no inherent connection between the investment horizon and the frequency with which portfolios can be rebalanced. Accordingly we now assume that the long-term investor can rebalance his portfolio every period; and we continue to use this assumption throughout the rest of the book.

Classic results of Samuelson (1969) and Merton (1969, 1971) give two sets of conditions under which the long-term investor acts myopically, choosing the same portfolio as a short-term investor. Portfolio choice will be myopic, first, if the investor has power utility and returns are IID. This result was originally derived using dynamic programming, but here we give a simple intuitive argument.

We start by noting that if returns are IID, no new information arrives between one period and the next so there is no reason for portfolio choice to change over time in a random fashion. We can therefore restrict attention to deterministic portfolio rules in which the risky asset share α_t may depend on time. We also note that K-period returns are lognormal if single-period returns are lognormal and IID. This means that the power-utility investor chooses a portfolio on the basis of the mean and variance of the K-period log portfolio return.

The K-period log return is just the sum of successive single-period log returns. For simplicity, consider an example in which $K = 2$. Then we can write

$$\begin{aligned} r_{p2,,t+2} - 2r_f &= (r_{p,t+1} - r_f) + (r_{p,t+2} - r_f) \\ &= \alpha_t(r_{t+1} - r_f) + \frac{1}{2}\alpha_t(1-\alpha_t)\sigma^2 \\ &\quad +\alpha_{t+1}(r_{t+2} - r_f) + \frac{1}{2}\alpha_{t+1}(1-\alpha_{t+1})\sigma^2, \end{aligned} \tag{2.30}$$

where we are allowing α_t and α_{t+1} to be different because the investor can freely rebalance his portfolio each period. The conditional variance of the 2-period log return is

$$\mathrm{Var}_t(r_{p,t,t+2}) = (\alpha_t^2 + \alpha_{t+1}^2)\sigma^2 \tag{2.31}$$

since both α_t and α_{t+1} are deterministic and hence are known at time t. The mean 2-period log return, adjusted by adding one-half the variance of the 2-period return, is

$$\mathrm{E}_t(r_{p,t,t+2}) + \frac{1}{2}\mathrm{Var}_t(r_{p,t,t+2}) = 2r_f + (\alpha_t + \alpha_{t+1})(\mathrm{E}r - r_f + \sigma^2/2). \tag{2.32}$$

The objective of a 2-period investor with power utility can be written as

$$\max \mathrm{E}_t(r_{p,t,t+2}) + \frac{1}{2}\mathrm{Var}_t(r_{p,t,t+2}) - \frac{\gamma}{2}\mathrm{Var}_t(r_{p,t,t+2}). \tag{2.33}$$

Thus the investor with $\gamma > 0$ always prefer a lower variance of log returns for a given variance-adjusted mean. But from (2.32), the variance-adjusted mean depends only on the sum $(\alpha_t+\alpha_{t+1})$. The investor can fix this sum and adjust the individual shares, α_t and α_{t+1}, to minimize the variance. Since variance depends on the sum of squares $(\alpha_t^2+\alpha_{t+1}^2)$, this is accomplished by setting $\alpha_t = \alpha_{t+1}$, a constant portfolio rule.

Once we know that the portfolio rule is constant, we also know that it must be the same as the optimal rule for a short-term investor. The reason is that in the last period before the horizon, the long-term investor has become a short-term investor and will choose the optimal short-term portfolio.

This argument for myopic portfolio choice straightforwardly extends to any long horizon K. It can be related to a puzzle posed by Mark Kritzman (2000) in his book *Puzzles of Finance*. In a chapter entitled "Half Stocks All the Time or All Stocks Half the Time?", Kritzman points out that these two strategies have the same expected simple return, but the latter strategy is riskier; thus a risk-averse investor should always prefer the former. This is precisely the effect that underlies the argument for a constant portfolio rule.

The second Samuelson-Merton condition for myopic portfolio choice is that the investor has log utility. In this case portfolio choice will be myopic even if asset returns are not IID. The argument here is particularly simple. Recall that the log utility investor chooses a portfolio that maximizes the expected log return. Equation (2.28) shows that the K-period log return is just the sum of 1-period log returns. Since the portfolio can be chosen freely each period, the sum is maximized by maximizing each of its elements separately, that is, by choosing each period the portfolio that is optimal for a 1-period log utility investor.

2.2.2 Fallacies of long-term portfolio choice

Although the conditions for myopic portfolio choice are simple, and their logic can be understood without resort to advanced mathematics, there has been much confusion about these issues over the years. One source of confusion is the common tendency to measure risk in units of standard deviation rather than variance, and to work with mean-standard deviation diagrams (like Figure 1.1) rather than mean-variance diagrams. With IID returns the variance of a cumulative risky return is proportional to the investment horizon K, but the standard deviation is proportional to the square root of K. (If returns are annualized, the variance is constant but the standard deviation shrinks in proportion to the square root of K.) Thus the Sharpe ratio of any risky investment, its mean excess return divided by its standard deviation, grows with the square root of K.

It is tempting to calculate long-horizon Sharpe ratios, observe that they are large, and conclude that lengthening the investment horizon somehow reduces risk in a manner analogous to the effect of diversifying a portfolio across uncorrelated risky assets. But this "time diversification" argument is a fallacy. Sharpe ratios cannot be compared across different investment horizons; they must always be measured over a common time interval.

A related fallacy is to argue that there is a single best long-term portfolio for all investors, regardless of their preferences. The proposed portfolio is the "growth-optimal" portfolio that maximizes the expected log return. As the investment horizon increases, this portfolio outperforms any other portfolio with higher and higher probability. In the limit, the probability that it outperforms goes to one. To understand this property of the growth-optimal portfolio, note that the difference between the cumulative growth-optimal log return and the log return on any other portfolio is a normally distributed random variable. Write the mean of this difference as $\mu(K)$ and the standard deviation as $\sigma(K)$. The mean $\mu(K)$ is positive because the growth-optimal log return has the highest mean of any available portfolio strategy. The probability that the difference is positive is $\Phi(-\mu(K)/\sigma(K))$, where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution. As the investment horizon K increases, the ratio $\mu(K)/\sigma(K)$ grows with $\sqrt{K}$, so $\Phi(-\mu(K)/\sigma(K))$ goes to one.

The fallacy is the claim that this property of the growth-optimal portfolio makes it the best portfolio for all long-term investors. It is, of course, the best portfolio for an investor with log utility; but investors with higher risk aversion should hold more conservative portfolios. Even though the growth-optimal portfolio will almost always outperform such conservative portfolios over long horizons, the loss when it does underperform is larger at long horizons, and this possibility is heavily weighted by conservative investors. This point was forcefully made by Samuelson (1979) in an article entitled "Why We Should Not Make Mean Log of Wealth Big Though Years to Act Are Long". The last paragraph of the article reads as follows: "No need to say more. I've made my point. And, save for the last word, have done so in prose of but one syllable." (p.306). A more recent popular discussion of the fallacy, which is easier to read because it allows itself a wider selection of words, is in Kritzman (2000, Chapter 3).

2.2.3 Power utility of consumption

The assumption that investors care only about wealth at a single horizon is analytically convenient but empirically troublesome. Most investors, whether they are individuals saving for retirement or institutions such as universities that live off endowment income, are concerned not with the level of wealth for its own sake, but with the standard of living that their wealth can support. In other words they consume out of wealth and derive utility from consumption rather than wealth.

In this section we assume that utility is defined over a stream of consumption. Once we make this assumption, the horizon plays a much smaller role in the analysis; it still defines a terminal date, but conditions at each intermediate date are also important. If the terminal date is distant, intermediate conditions dominate the solution, which will not depend sensitively on the exact choice of terminal date. In fact, we can let the terminal date go to infinity and work with an attractively simple infinite-horizon model. We can vary the effective investment horizon by varying the time discount factor that determines the relative weights investors place on the near-term future versus the distant future. This is our mode of analysis for most of the rest of the book.

We first assume that investors have time-separable power utility, defined over consumption:

$$\max \mathrm{E}_t \sum_{i=0}^{\infty} \delta^i U(C_{t+i}) = \mathrm{E}_t \sum_{i=0}^{\infty} \delta^i \frac{C_{t+i}^{1-\gamma} - 1}{1-\gamma}. \tag{2.34}$$

Here δ is the time discount factor. When δ is large, investors place relatively high weight on the distant future. As δ shrinks, they place more and more weight on the near future; in the limit as δ approaches zero, they behave like single-period investors. Investors face the intertemporal budget constraint that wealth next period equals the portfolio return times reinvested wealth, that is, wealth today less what is subtracted for consumption:

$$W_{t+1} = (1 + R_{p,t+1})(W_t - C_t). \tag{2.35}$$

This objective function and budget constraint imply the following first-order condition or *Euler equation* for optimal consumption choice:

$$U'(C_t) = \mathrm{E}_t[\delta U'(C_{t+1})(1 + R_{i,t+1})], \tag{2.36}$$

where $(1 + R_{i,t+1})$ denotes any available return, for example the riskless return $(1 + R_{f,t+1})$, the risky return $(1 + R_{t+1})$, or the portfolio return $(1 + R_{p,t+1})$. Equation (2.36) says that at the optimum, the marginal cost of saving an extra dollar for one period must equal the marginal benefit. The marginal cost is the marginal utility of a dollar of consumption, $U'(C_t)$. The marginal benefit is the expectation of the payoff if the dollar is invested in an available asset for one period, $(1 + R_{i,t+1})$, times the marginal utility of an extra dollar of consumption next period, $U'(C_{t+1})$, discounted back to the present at rate δ.

We can divide (2.36) through by $U'(C_t)$ and use the power utility condition that $U'(C_t) = C_t^{-\gamma}$ to rewrite as

$$1 = \mathrm{E}_t \left[\delta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} (1 + R_{i,t+1}) \right]. \tag{2.37}$$

In the case where the return is riskless, $R_{i,t+1} = R_{f,t+1}$, it is known at time t and thus can be brought outside the expectations operator; we have

$$\frac{1}{(1 + R_{f,t})} = \mathrm{E}_t \left[\delta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} \right]. \tag{2.38}$$

The term $\delta(C_{t+1}/C_t)^{-\gamma}$ that appears in all these expressions is known as the "stochastic discount factor" or SDF because it can be used to discount expected payoffs on any assets to find their prices. The consumption of any investor who is able to freely trade in the financial markets can be used as an SDF. The field of asset pricing asks how the properties of the SDF explain the properties of asset returns, and how the SDF is determined by the general equilibrium of the economy; in this book we take a partial-equilibrium perspective and seek to explain the properties of optimal consumption and portfolio choice given asset returns. Campbell (2000) and Cochrane (2000) survey the SDF methodology.

A lognormal consumption-based model

The natural way to proceed, given our earlier discussion, is to assume that asset returns and consumption are jointly lognormal, and to work with loglinear versions of these equations. Hansen and Singleton (1983) pioneered this approach. The log form of the riskless-rate Euler equation (2.38) can be written as

$$\mathrm{E}_t[\Delta c_{t+1}] = \frac{\log \delta}{\gamma} + \frac{1}{\gamma} r_{f,t+1} + \frac{\gamma}{2} \sigma_{ct}^2. \tag{2.39}$$

The three terms on the right-hand side of (2.39) correspond to three forces acting on consumption. First, a patient investor with a high time discount factor δ is inherently willing to postpone consumption. Second, a high interest rate gives an investor an incentive to postpone consumption. Postponing consumption means raising consumption in the future relative to consumption today, tilting the consumption path upwards; however diminishing marginal utility of consumption limits the investor's willingness to tolerate any deviation from a flat consumption path. The investor's willingness to tilt consumption in response to incentives is known as the *elasticity* *of intertemporal substitution in consumption* (EIS). In a model with power utility, the EIS equals the reciprocal of risk aversion, $1/\gamma$. Thus with a high γ the investor is extremely unwilling to tilt consumption and the planned consumption growth rate will change only slightly with the time discount factor and the riskless interest rate. The third term on the right-hand side of (2.39) represents the effect of uncertainty. A risk-averse investor will respond to uncertainty by increasing precautionary saving, again tilting the consumption path upwards.

It is also possible to write expected consumption growth in terms of the expected portfolio return. The first term on the right-hand side of (2.39) is unchanged by doing this, the second term becomes $(1/\gamma)E_t r_{p,t+1}$, and the third term involves the variances and covariances of the portfolio return and consumption growth. This alternative representation is more convenient for some purposes, and will be used in our analysis of Epstein-Zin utility in the next section.

The Euler equation for power utility can also be used to describe the risk premium on a single risky asset over the riskless interest rate. The log form of the general Euler equation (2.37), less γ times (2.39), is

$$\mathrm{E}_t r_{t+1} - r_{f,t+1} + \frac{\sigma_t^2}{2} = \gamma \mathrm{Cov}_t(r_{t+1,} \Delta c_{t+1}). \tag{2.40}$$

This says that in equilibrium, the expected excess return on the risky asset must equal risk aversion γ times the covariance of the asset return with consumption growth. A similar equation describes each risky asset's risk premium in a model with multiple risky assets. In asset pricing theory, this equation is used to explain assets' risk premia and is known as the *consumption capital asset pricing model* or CCAPM; here we take the risky asset return as given and seek a consumption rule and portfolio strategy that will make (2.40) hold.

A constant consumption-wealth ratio

A difficulty with the lognormal consumption-based model is that the intertemporal budget constraint (2.35) is not generally loglinear, because consumption is *subtracted* from wealth before *multiplying* by the portfolio return. The combination of subtraction and multiplication creates an intractable nonlinearity.

In later chapters of this book, following Campbell (1993), we will approach this problem by approximating the budget constraint. For now, however, we assume that the consumption-wealth ratio is constant, in which case the budget constraint becomes loglinear. We solve the model under this assumption, and then seek to find conditions that justify the assumption. A constant consumption-wealth ratio can be written as

$$\frac{C_t}{W_t} = b, \tag{2.41}$$

and the budget constraint (2.35) can then be written in log form as

$$\begin{aligned}\Delta w_{t+1} &= r_{p,t+1} + \log(1-b) \\ &= r_{f,t+1} + \alpha_t(r_{t+1} - r_{f,t+1}) \\ &\quad + \frac{1}{2}\alpha_t(1-\alpha_t)\sigma_t^2 + \log(1-b),\end{aligned} \tag{2.42}$$

where the second equality substitutes in from (2.21).

The constant consumption-wealth ratio (2.41) also implies that the growth rate of consumption equals the growth rate of wealth, so the terms in consumption in (2.39) and (2.40) can be rewritten in terms of wealth. The formula for a single risky asset's expected excess return, (2.40), becomes

$$\mathrm{E}_t r_{t+1} - r_{f,t+1} + \sigma_t^2/2 = \gamma \mathrm{Cov}_t(r_{t+1}, \Delta w_{t+1}) = \gamma \alpha_t \sigma_t^2, \tag{2.43}$$

where the second equality follows from (2.42). Solving this equation for α_t, we once again obtain the myopic solution (2.24):

$$\alpha_t = \frac{\mathrm{E}_t r_{t+1} - r_{f,t+1} + \sigma_t^2/2}{\gamma \sigma_t^2}.$$

It is straightforward to show that the myopic solution for multiple risky assets, (2.26), is equally valid for a long-term investor with a constant consumption-wealth ratio.

In a model with power utility, then, the assumption that the consumption-wealth ratio is constant immediately leads to the conclusion that portfolio choice is myopic. Under what conditions can this assumption be justified? The answer to this question is by now familiar. First, if returns are IID, there are no changes over time in investment opportunities that might induce changes in consumption relative to wealth. The scale-independence of the power utility function implies that consumption is a constant fraction of wealth in this case. Second, if risk aversion $\gamma = 1$, consumption is again a constant fraction of wealth. The intuition for this result is that changing investment opportunities have opposing effects on consumption relative to wealth. An improved investment opportunity, say a higher riskless interest rate, raises the amount that can be consumed each period without depleting wealth. This *income effect* tends to increase consumption relative to wealth. On the other hand, an improved investment opportunity creates an incentive to postpone consumption to the future. This *substitution effect* tends to decrease consumption relative to wealth. In the power-utility model, the log-utility case $\gamma = 1$ is the case where income and substitution effects exactly cancel out, so that the consumption-wealth ratio is always constant regardless of any fluctuations in investment opportunities.

To prove these statements, one can use the budget constraint (2.42) to find the conditional mean and variance of wealth, and substitute into (2.39). The resulting equation can be solved for a constant b if asset returns are IID or if $\gamma = 1$. Intuitively, (2.39) says that expected consumption growth should move $1/\gamma$ for one with the expected return. This is consistent with a constant consumption-wealth ratio if the expected return is constant, or if expected consumption growth adjusts one for one with the expected return so that the desired changes in consumption growth can be financed just by the variation in the expected return itself, without any need for savings adjustments.

2.2.4 Epstein-Zin utility

Despite the many attractive features of the power-utility model, it does have one highly restrictive feature. Power utility implies that the consumer's elasticity of intertemporal substitution, ψ, is the reciprocal of the coefficient of relative risk aversion, γ. Yet it is not clear that these two concepts should be linked so tightly. Risk aversion describes the consumer's reluctance to substitute consumption across states of the world and is meaningful even in an atemporal setting, whereas the elasticity of intertemporal substitution describes the consumer's willingness to substitute consumption over time and is meaningful even in a deterministic setting. Epstein and Zin (1989, 1991) and Weil (1989) use the theoretical framework of Kreps and Porteus (1978) to develop a more flexible version of the basic power utility model.The Epstein-Zin model retains the desirable scale-independence of power utility but breaks the link between the parameters γ and ψ.

The Epstein-Zin objective function is defined recursively by

$$U_t = \left\{ (1-\delta) C_t^{\frac{1-\gamma}{\theta}} + \delta \left(\mathrm{E}_t \, U_{t+1}^{1-\gamma} \right)^{\frac{1}{\theta}} \right\}^{\frac{\theta}{1-\gamma}}, \tag{2.44}$$

where $\theta \equiv (1-\gamma)/(1-1/\psi)$. When $\gamma = 1/\psi$, $\theta = 1$ and the recursion

(2.44) becomes linear; it can then be solved forward to yield the familiar time-separable power utility model.

The Epstein-Zin model can be understood by reference to Figure 2.2. The horizontal axis shows the elasticity of intertemporal substitution, ψ, while the vertical axis shows the coefficient of relative risk aversion, γ. The set of points with unit elasticity of intertemporal substitution is drawn as a vertical line, while the set of points with unit relative risk aversion is drawn as a horizontal line. The set of points with power utility is drawn as the hyperbola $\gamma = 1/\psi$. Log utility is the point where all three lines cross; it has $\gamma = \psi = 1$.

The nonlinear recursion (2.44) does not look at all easy to work with. Fortunately Epstein and Zin have shown, using dynamic programming arguments, that if the intertemporal budget constraint takes the form (2.35) (that is, if the investor finances consumption entirely from financial wealth and does not receive labor income), then there is an Euler equation of the form

$$1 = \mathrm{E}_t \left[\left\{ \delta \left(\frac{C_{t+1}}{C_t} \right)^{-\frac{1}{\psi}} \right\}^{\theta} \left\{ \frac{1}{(1+R_{p,t+1})} \right\}^{1-\theta} (1+R_{i,t+1}) \right], \qquad (2.45)$$

where as before $(1 + R_{i,t+1})$ is the gross return on any available asset, including the riskless asset and the portfolio itself.

Equation (2.45) simplifies somewhat if we set $(1+R_{i,t+1}) = (1+R_{p,t+1})$. If the portfolio return and consumption are jointly lognormal, we then find that expected consumption growth equals

$$\mathrm{E}_t[\Delta c_{t+1}] = \psi \log \delta + \psi \mathrm{E}_t r_{p,t+1} + \frac{\theta}{2\psi} \mathrm{Var}_t[\Delta c_{t+1} - \psi r_{p,t+1}]. \qquad (2.46)$$

Expected consumption growth is determined by time preference, the expected portfolio return, and the effects of uncertainty summarized in the variance term. Note that the elasticity of intertemporal substitution ψ, and not the coefficient of relative risk aversion γ, determines the response of expected consumption growth to variations in the expected return. Randomness in future consumption growth, relative to portfolio returns, increases precautionary savings and lowers current consumption if $\theta > 0$ (a condition satisfied by power utility for which $\theta = 1$), but reduces precautionary savings and increases current consumption if $\theta < 0$.

When there is a single risky asset, the premium on the risky asset over the safe asset is

$$\mathrm{E}_t r_{t+1} - r_{f,t+1} + \frac{\sigma_t^2}{2} = \theta \frac{\mathrm{Cov}_t(r_{t+1}, \Delta c_{t+1})}{\psi} + (1-\theta)\mathrm{Cov}_t(r_{t+1}, r_{p,t+1}). \quad (2.47)$$

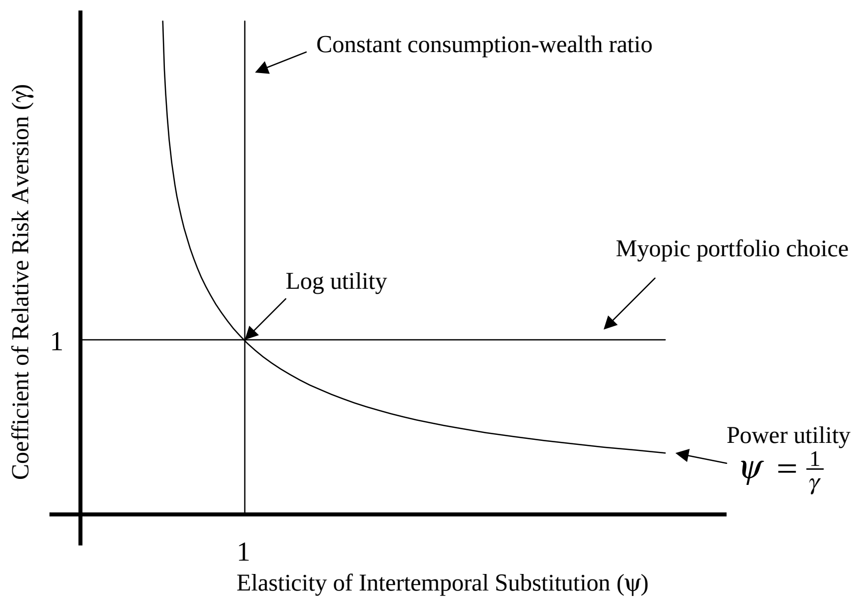

Figure 2.2: Epstein-Zin utility

The expected excess return on the risky asset is a weighted average of the risky asset's covariance with consumption growth (divided by the elasticity of intertemporal substitution ψ) and asset i's covariance with the portfolio return. The weights are θ and $1-\theta$ respectively.

A similar equation holds for each risky asset in a model with multiple risky assets and no short-term riskless asset. In vector notation, we have

$$\mathrm{E}_t(\mathbf{r}_{t+1} - r_{0,t+1}\boldsymbol{\iota}) + \frac{\boldsymbol{\sigma}_t^2}{2} = \frac{\theta}{\psi}\boldsymbol{\sigma}_{ct} + (1-\theta)\boldsymbol{\sigma}_{pt} - \boldsymbol{\sigma}_{0t}, \tag{2.48}$$

where as before $r_{0,t+1}$ is the total return on the risky benchmark asset, $\boldsymbol{\sigma}_t^2$ is the vector of excess-return variances, $\boldsymbol{\sigma}_{ct}$ is the vector of excess-return covariances with consumption growth, $\boldsymbol{\sigma}_{pt}$ is the vector of excess-return covariances with the total return on the portfolio, and $\boldsymbol{\sigma}_{0t}$ is the vector of excess-return covariances with the total return on the benchmark asset. In an asset pricing context, this equation explains any asset's risk premium by reference to both its consumption covariance (the consumption CAPM) and its covariance with the investor's overall portfolio (the traditional CAPM). In the power utility case, $\theta = 1$ and we have a pure consumption CAPM.

The familiar conditions for myopic portfolio choice follow immediately from (2.47). If asset returns are IID, then consumption is a constant fraction of wealth and covariance with consumption growth equals covariance with portfolio return. In this case the right-hand side of (2.47) can be rewritten as $(\theta/\psi + 1 - \theta)\mathrm{Cov}_t(r_{t+1}, r_{p,t+1}) = \gamma \mathrm{Cov}_t(r_{t+1}, r_{p,t+1})$, which implies the myopic portfolio rule. If relative risk aversion $\gamma = 1$, then $\theta = 0$ and the right-hand side of (2.47) is just $\mathrm{Cov}_t(r_{t+1}, r_{p,t+1})$, which again implies the myopic portfolio rule. This derivation makes it clear that what is required for myopic portfolio choice is unit relative risk aversion, not a unit elasticity of intertemporal substitution. Log utility is the special case where both relative risk aversion and the elasticity of intertemporal substitution (EIS) equal one.

The case of a unit EIS requires careful handling in this model. As ψ approaches one, θ approaches positive or negative infinity. Equation (2.46) can only be satisfied if $\mathrm{Var}_t[\Delta c_{t+1} - r_{p,t+1}] = 0$, which implies a constant consumption-wealth ratio. To analyze portfolio choice in this case, one must take appropriate limits of the terms on the right-hand side of (2.47). Giovannini and Weil (1989) have done this analysis and have shown that the model with $\psi = 1$ does not have myopic portfolio choice unless $\gamma = 1$ (in which case we have log utility). The constancy of the consumption-wealth ratio in the unit EIS model makes this a particularly tractable specification. (This is also true in a continuous-time setting, as shown by Schroder and

Skiadas 1999). In the succeeding chapters we will use the unit-EIS case as a benchmark case in which we can get exact solutions, and around which we can construct approximate solutions.

Finally, Epstein and Zin have also shown that the maximized utility function or value function per unit wealth, $V_t \equiv U_t/W_t$, is related to consumption per unit wealth C_t/W_t by the expression

$$V_t = (1-\delta)^{-\frac{\psi}{1-\psi}} \left(\frac{C_t}{W_t}\right)^{\frac{1}{1-\psi}}. \tag{2.49}$$

Two special cases are worth noting. First, as ψ approaches one, the exponents in (2.49) increase without limit. The value function has a finite limit, however, because the ratio C_t/W_t approaches $(1-\delta)$ as shown by Giovannini and Weil (1989). Second, as ψ approaches zero, V_t approaches C_t/W_t. A consumer who is extremely reluctant to substitute intertemporally consumes the annuity value of wealth each period, and this consumer's utility per dollar is the annuity value of the dollar.

2.3 Conclusion

Does the investment horizon affect portfolio choice? In this chapter we have shown that it may not. We have assumed that investors' relative risk aversion does not depend systematically on their wealth, an assumption that is required to explain the stability of interest rates and asset returns through two centuries of economic growth. Under this assumption the investment horizon is irrelevant for investors who have only financial wealth and who face constant investment opportunities. Even if investment opportunities are time-varying, the investment horizon is still irrelevant for investors whose relative risk aversion equals one. Such investors should behave myopically, choosing the portfolio that has the best short-term characteristics. Popular arguments to the contrary, such as the claim that long-term investors can afford to take greater risk because they have "time to ride out the ups and downs of the market", are simply wrong under these conditions.

Legitimate arguments for horizon effects on portfolio choice depend on violations of the conditions for myopic portfolio choice discussed in this chapter. In our view there is strong empirical evidence that these conditions fail in various ways. The rest of the book is devoted to an exploration of portfolio choice in the presence of such failures.

Chapter 3

Who Should Buy Long-Term Bonds?

In Chapter 2 we explored in detail the conditions for myopic portfolio choice. We now begin to develop a theory of long-term portfolio choice when these conditions fail. To keep things as simple as possible, we assume initially that all the random variables that are relevant for investors are lognormally distributed with constant variances and covariances. This assumption is not innocuous in the context of portfolio choice theory, because it requires both that the investor's portfolio return has a constant variance, and that the returns on individual assets have constant variances and a constant covariance with the portfolio. These conditions are consistent with one another only if the composition of the portfolio is constant, which in turn is optimal—given constant variances—only if the expected excess returns on all assets are constant. Asset returns can change over time, but they must move in parallel with the riskless interest rate. In other words, our initial model allows only a limited form of time-variation in investment opportunities, driven by movements in the short-term interest rate. In the next chapter we generalize the model to allow movements in risk premia as well.

Throughout this and the next chapter we consider investors who have only financial wealth and no labor income, deferring a consideration of labor income to Chapter 6. We assume that investors care not about wealth for its own sake, but about the consumption stream that can be financed by wealth. To keep the analysis simple we assume that the investor is infinitely lived; we can vary the effective investment horizon by varying the investor's rate of time preference and thus varying the relative importance of the near future and the distant future. We assume that the investor has the Epstein-Zin preferences developed in section 2.2.4, with constant relative risk aversion γ and a constant elasticity of intertemporal substitution ψ that need not be related to one another.

The key problem in solving an intertemporal model of this sort is that the investor's intertemporal budget constraint is nonlinear. In section 3.1.1 we approach this problem by taking a loglinear approximation of the budget constraint, as first proposed by Campbell (1993). The approximation is exact if the consumption-wealth ratio is constant (as it will be if $\psi = 1$), and it is accurate if the consumption-wealth ratio is not too variable (as it will be if ψ is not too far from one). In section 3.1.2 we show how the approximate budget constraint can be used to substitute consumption out of the Euler equations of the Epstein-Zin model, giving an expression relating assets' risk premia to their covariances with current portfolio returns and revisions in expected future portfolio returns.

Section 3.1.3 applies these methods to the portfolio problem, deriving an explicit expression for the portfolio weight on a single risky asset. When the investor has risk aversion greater than one, the demand for the risky asset is affected not only by the asset's risk premium in relation to its variance, but also by its covariance with revisions in expected future interest rates. An asset whose value increases when interest rates fall is a desirable hedge against declines in interest rates that would otherwise reduce the income thrown off by the portfolio. This is the intertemporal hedging effect first emphasized by Merton (1973). Section 3.1.4 develops this idea further. Long-term bond prices move inversely with interest rates, so they are good intertemporal hedges. As the investor's risk aversion increases, the optimal portfolio approaches an inflation-indexed perpetuity that pays one unit of real consumption forever. In an important sense this asset is the riskless asset for a long-term investor; even though it may have an unstable capital value in the short term, it finances a riskless consumption stream over the long term. Section 3.1.5 generalizes the analysis to the case where there are multiple risky assets and there may be no short-term riskless asset. This also allows us to find the optimal portfolio even when the investor faces borrowing and short-sales constraints.

In section 3.2 we develop a more specific model and fit it to historical interest-rate data from the United States (and the United Kingdom, to be added later). We specify the model in section 3.2.1, present US estimates in section 3.2.2, and derive the implied optimal portfolios in section 3.2.3. Section 3.3 concludes.

3.1 Long-Term Portfolio Choice in a Model with Constant Variances and Risk Premia

3.1.1 Approximation of the intertemporal budget constraint

Recall that the intertemporal budget constraint, in a model with consumption at every date, is

$$W_{t+1} = (1 + R_{p,t+1})(W_t - C_t). \tag{3.1}$$

A central problem in the theory of intertemporal portfolio choice is that this budget constraint is nonlinear because consumption is subtracted from wealth before the portfolio return multiplies the remainder. In other words, only reinvested wealth earns the portfolio return and not all wealth is reinvested.

In the last chapter we concentrated on models in which the consumption-wealth ratio is constant. In this case reinvested wealth $(W_t - C_t)$ is a constant fraction of total wealth, and (3.1) can be rewritten in loglinear form.

We now develop an alternative approach to the problem of nonlinearity. Following Campbell (1993), we approximate the budget constraint around the mean of the consumption-wealth ratio. We first divide (3.1) by W_t to get

$$\frac{W_{t+1}}{W_t} = (1 + R_{p,t+1})(1 - \frac{C_t}{W_t}). \tag{3.2}$$

Taking logs, this becomes

$$\Delta w_{t+1} = r_{p,t+1} + \log(1 - \exp(c_t - w_t)). \tag{3.3}$$

The second term on the right-hand side of (3.3) is a nonlinear function of the log consumption-wealth ratio. If that ratio is not too variable, this can be well approximated using a first-order Taylor expansion around its mean. Details are given in the Appendix; the resulting expression is

$$\Delta w_{t+1} = k + r_{p,t+1} + \left(1 - \frac{1}{\rho}\right)(c_t - w_t), \tag{3.4}$$

where k and ρ are parameters of linearization. The parameter ρ is defined by $\rho = 1 - \exp(\overline{c - w})$. When the consumption-wealth ratio is constant, then ρ can be interpreted as $(W - C)/W$, the ratio of reinvested wealth to total wealth. The parameter k is given by the messy expression $k = \log(\rho) + (1 - \rho)\log(1 - \rho)/\rho$.

It is helpful to derive a long-term version of this budget constraint. To do this we can use the trivial equality

$$\Delta w_{t+1} = \Delta c_{t+1} + (c_t - w_t) - (c_{t+1} - w_{t+1}). \tag{3.5}$$

Equating the left-hand sides of (3.4) and (3.5), we obtain a difference equation in the log consumption-wealth ratio. We can solve forward, assuming that $\lim_{j \longrightarrow \infty} \rho^j (c_{t+j} - w_{t+j}) = 0$ (a condition that will hold, for example, if the consumption-wealth ratio is stationary), to get

$$c_t - w_t = \sum_{j=1}^{\infty} \rho^j (r_{p,t+j} - \Delta c_{t+j}) + \frac{\rho k}{1-\rho}. \tag{3.6}$$

This equation says that a high consumption-wealth ratio today must be followed either by high returns on invested wealth or by low consumption growth. That is, high consumption today will deplete wealth and hence future consumption possibilities unless it is offset by high investment returns. This follows simply from the intertemporal budget constraint; there is no model of optimal behavior in (3.6).

Equation (3.6) holds ex post, but it also holds ex ante; if one takes expectations of (3.6) at time t, the left-hand side is unchanged, and the right-hand side becomes an expected discounted value:

$$c_t - w_t = \mathrm{E}_t \sum_{j=1}^{\infty} \rho^j (r_{p,t+j} - \Delta c_{t+j}) + \frac{\rho k}{1-\rho}. \tag{3.7}$$

Equation (3.7) can be substituted into (3.4) and (3.5) to obtain

$$\begin{aligned} c_{t+1} - \mathrm{E}_t c_{t+1} &= (\mathrm{E}_{t+1} - \mathrm{E}_t) \sum_{j=0}^{\infty} \rho^j r_{p,t+1+j} \\ &\quad - (\mathrm{E}_{t+1} - \mathrm{E}_t) \sum_{j=0}^{\infty} \rho^j \Delta c_{t+1+j}. \end{aligned} \tag{3.8}$$

Equation (3.8) says that an upward surprise in consumption today must correspond to an unexpected return on wealth today (the first term in the first sum on the right-hand side of the equation), or to news that future returns will be higher (the remaining terms in the first sum), or to a downward revision in expected future consumption growth (the second sum on the right-hand side).

These formulas are analogous to expressions developed by Campbell and Shiller (1988) and Campbell (1991) relating dividends, stock prices, and stock returns. Wealth can be thought of as an asset that pays consumption as its dividend. This insight applies both to an individual's wealth and consumption, and to aggregate wealth and consumption; thus the assumption commonly made in empirical finance research, that an aggregate stock index is a good proxy for the market portfolio of aggregate wealth, is equivalent to the assumption of Lucas (1978) and Mehra and Prescott (1985) that stocks are priced as if they pay dividends equal to the aggregate consumption of the economy.

3.1.2 Substituting out consumption

We have shown that the consumption-wealth ratio can be related to expected future returns and consumption growth. The next step is to substitute expected consumption growth out of the model. For this purpose we can use the Euler equation for Epstein-Zin utility under lognormality, (2.46):

$$\mathrm{E}_t[\Delta c_{t+1}] = \psi \log \delta + \psi \mathrm{E}_t r_{p,t+1} + \frac{\theta}{2\psi} \mathrm{Var}_t[\Delta c_{t+1} - \psi r_{p,t+1}].$$

If consumption and the portfolio return are not only lognormal but homoskedastic, then the variance term in (2.46) is constant and we can rewrite as

$$\mathrm{E}_t[\Delta c_{t+1}] = \mu + \psi \mathrm{E}_t r_{p,t+1}, \tag{3.9}$$

where the intercept μ includes not only the pure rate of time preference but also the effects of risk on consumption.

Substituting (3.9) into (3.7), we find

$$c_t - w_t = (1-\psi)\mathrm{E}_t \sum_{j=1}^{\infty} \rho^j r_{p,t+j} + \frac{\rho(k-\mu)}{1-\rho}. \tag{3.10}$$

The log consumption-wealth ratio depends on the expected discounted value of all future portfolio returns, with positive sign if $\psi < 1$ and negative sign if $\psi > 1$. There are opposing income and substitution effects of an increased portfolio return. On the one hand, if the portfolio return is higher, consumption can be higher in all periods for any value of wealth; this is the positive income effect of the portfolio return on consumption. On the other hand, if the portfolio return is higher, there is a greater incentive to delay consumption, cutting consumption today in order to exploit more favorable investment opportunities; this is the negative substitution effect of the portfolio return on consumption. The income effect dominates if $\psi < 1$ and the substitution effect dominates if $\psi > 1$. If $\psi = 1$, the two effects cancel and the consumption-wealth ratio is constant as previously noted.

Equation (3.10) allows us to rewrite (3.8) as

$$c_{t+1} - \mathrm{E}_t c_{t+1} = r_{p,t+1} - \mathrm{E}_t r_{p,t+1} + (1-\psi)(\mathrm{E}_{t+1} - \mathrm{E}_t)\sum_{j=0}^{\infty} \rho^j r_{p,t+1+j}. \quad (3.11)$$

The innovation in consumption is the surprise component of the portfolio return, which has a one-for-one effect because of the scale-independence of the utility function, plus $(1-\psi)$ times the revision in expectations of future returns. If $\psi < 1$, a positive surprise about future returns increases consumption today through the dominant income effect; if $\psi > 1$, a positive surprise about future returns reduces consumption today through the dominant substitution effect.

Earlier we assumed that consumption and asset returns are jointly lognormal and homoskedastic. We can now see what is required to justify this assumption. According to (3.11), we need that log portfolio returns and revisions in expectations of future log portfolio returns are normal and homoskedastic, as they will be for example in a linear time-series model for log returns and other state variables.

Recall that under Epstein-Zin utility with a single risky asset, the premium on the risky asset over the safe asset is given by (2.47):

$$\mathrm{E}_t r_{t+1} - r_{f,t+1} + \frac{\sigma_t^2}{2} = \theta \frac{\mathrm{Cov}_t(r_{t+1}, \Delta c_{t+1})}{\psi} + (1-\theta)\mathrm{Cov}_t(r_{t+1}, r_{p,t+1}).$$

Although we are assuming that conditional variances and covariances are constant over time, we retain their time subscripts to make it clear that they are calculated conditional on time t information. Equation (3.11) implies that the covariance with consumption that appears in this expression can be replaced by the covariance with the portfolio return, plus $(1-\psi)$ times the covariance with revised expectations about future returns. Using the relation between the parameters θ, γ, and ψ, we find that

$$\begin{aligned} \mathrm{E}_t r_{t+1} - r_{f,t+1} + \frac{\sigma_t^2}{2} &= \gamma \mathrm{Cov}_t(r_{t+1}, r_{p,t+1}) \quad (3.12) \\ &\quad + (\gamma - 1)\mathrm{Cov}_t(r_{t+1}, (\mathrm{E}_{t+1} - \mathrm{E}_t)\sum_{j=0}^{\infty} \rho^j r_{p,t+1+j}). \end{aligned}$$

Campbell (1993) derived this equation, and Campbell (1996) used it in an asset pricing context, asking whether the CAPM, as extended by the second term on the right-hand side of (3.12), could explain the pattern of risk premia in US financial markets.

This analysis shows that under Epstein-Zin utility, there is an elegant separation between the elasticity of intertemporal substitution ψ and the coefficient of relative risk aversion γ. Given the loglinearization coefficient ρ, only the parameter ψ appears in equations (3.10) and (3.11) that relates consumption to returns, and only the parameter γ appears in the equation (3.12) that relates the risk premium to the second moments of asset returns. A caveat is that in general the loglinearization parameter ρ itself depends on both ψ and γ; however we shall see that this dependence is empirically weak and does not seriously undermine the separation of ψ and γ.

3.1.3 Application to portfolio choice

So far we have assumed only that consumption and the optimal portfolio are jointly lognormal with constant variances. In order to apply (3.12) in a model of portfolio choice, we need to make the further assumption that the available individual assets have constant variances and risk premia. This implies that variation in the expected portfolio return is entirely due to variation in the riskless interest rate, so the revisions in expected future portfolio returns in (3.12) are equivalent to revisions in expected future riskless interest rates: $(\mathrm{E}_{t+1}-\mathrm{E}_t)\sum_{j=0}^{\infty}\rho^j r_{p,t+1+j} = (\mathrm{E}_{t+1}-\mathrm{E}_t)\sum_{j=0}^{\infty}\rho^j r_{f,t+1+j}$. It also implies that optimal portfolio weights are constant, reconciling the assumptions of constant variances for both individual assets and the portfolio return.

In a model with a single risky asset, we have $\mathrm{Cov}_t(r_{t+1}, r_{p,t+1}) = \alpha_t\sigma_t^2$. Substituting into (3.12) and rearranging, we find that the optimal portfolio weight on the risky asset is a constant α given by

$$\begin{aligned}\alpha &= \frac{1}{\gamma}\frac{\mathrm{E}_t r_{t+1} - r_{f,t+1} + \sigma_t^2/2}{\sigma_t^2} \\ &\quad + \left(1-\frac{1}{\gamma}\right)\frac{\mathrm{Cov}_t(r_{t+1}, -(\mathrm{E}_{t+1}-\mathrm{E}_t)\sum_{j=0}^{\infty}\rho^j r_{f,t+1+j})}{\sigma_t^2}. \quad (3.13)\end{aligned}$$

The demand for the risky asset is a weighted average of two desirable attributes. The first attribute is the asset's risk premium, relative to its variance, and the second, intertemporal attribute is the asset's covariance with reductions in expected future interest rates, again relative to its variance. The weight on the first attribute is relative risk tolerance $(1/\gamma)$, which becomes negligible as γ increases; a highly conservative investor does not buy the risky asset for its risk premium. The weight on the second attribute is $1 - (1/\gamma)$, which is zero for a myopic investor with $\gamma = 1$ but approaches 1 as γ increases. A highly conservative investor holds the risky asset only if it covaries with declines in interest rates, compensating the portfolio for the reduction in income that occurs when interest rates fall.

A variant of (3.13) was derived by Restoy (1992). Using (3.10), Restoy noted that revisions in expected future portfolio returns are proportional to surprises in the log consumption-wealth ratio, $(\mathrm{E}_{t+1}-\mathrm{E}_t)\sum_{j=0}^{\infty}\rho^j r_{p,t+1+j} = (\mathrm{E}_{t+1}-\mathrm{E}_t)(c_{t+1} - w_{t+1})/(1-\psi)$. Hence we can write

$$\alpha = \frac{1}{\gamma}\frac{\mathrm{E}_t r_{t+1} - r_{f,t+1} + \sigma_t^2/2}{\sigma_t^2} + \left(1 - \frac{1}{\gamma}\right)\left(\frac{1}{1-\psi}\right)\frac{\mathrm{Cov}_t(r_{t+1}, -(c_{t+1} - w_{t+1}))}{\sigma_t^2}. \quad (3.14)$$

This equation shows that the intertemporal component of asset demand works through covariance of the risky asset with the consumption-wealth ratio. However it can be misleading because it suggests that the parameter ψ plays a role in asset demand whereas in fact ψ cancels out of the previous equation.

A third transformation of (3.13) uses the Epstein-Zin result (2.49) relating the value function per unit wealth to the consumption-wealth ratio. Taking logs of (2.49), the second term in (3.14) can be restated in terms of the covariance of the risky asset return with the value function,

$$\alpha = \frac{1}{\gamma}\frac{\mathrm{E}_t r_{t+1} - r_{f,t+1} + \sigma_t^2/2}{\sigma_t^2} + \left(1 - \frac{1}{\gamma}\right)\frac{\mathrm{Cov}_t(r_{t+1}, -v_{t+1})}{\sigma_t^2}. \quad (3.15)$$

This shows that the intertemporal component of asset demand is determined by the covariance of the risky asset return with the investor's utility per unit wealth, which varies over time with investment opportunities.

One can use these three equations to understand the intellectual history of research on long-term portfolio choice. Merton (1971, 1973) introduced the concept of intertemporal hedging demand for risky assets. He worked with the indirect utility or value function defined over wealth and state variables; thus his approach was similar in spirit to (3.15). Breeden (1979) first used covariance with consumption as a measure of risk for a long-term investor; thus his approach was close to (3.14). Neither author derived an explicit solution comparable to (3.13), relating portfolio demands to covariances of assets with exogenous variables. Equation (3.13) depends on specific assumptions, but it can be used to obtain general insights. Note in particular that intertemporal hedging demand in (3.13) depends on the discounted present value of all future interest rates; thus portfolio choice of long-term investors is affected far more strongly by persistent variations in investment opportunities than by transitory variations.

3.1.4 What is the riskless asset?

The identity of the riskless asset is a fundamental issue in finance. It is conventional to think of the riskless asset as an asset that has a stable capital value in the short term, such as a Treasury bill or money market fund. Such an asset has a known return over one period, and will be held by an infinitely conservative short-term investor.

For a long-term investor, however, a strategy of rolling over Treasury bills is not necessarily safe because maturing bills must be reinvested at unknown future real interest rates. Over thirty years ago Modigliani and Sutch (1966) made is point particularly clearly, writing

> Suppose a person has an n period habitat; that is, he has funds which he will not need for n periods and which, therefore, he intends to keep invested in bonds for n periods. If he invests in n period bonds, he will know exactly the outcome of his investments as measured by the terminal value of his wealth.... If, however, he stays short, his outcome is uncertain.... Thus, if he has risk aversion, he will prefer to stay long (pp. 183–184).

Modigliani and Sutch's assumption, that an investor cares about only about wealth at a single future date, is somewhat artificial. However a similar point applies to a long-term investor of the sort modelled in this book, who cares about the stream of consumption or standard of living that can be supported by wealth.

To show this, we consider what happens to portfolio choice as the coefficient of relative risk aversion increases. In this case relative risk tolerance $(1/\gamma)$ goes to zero so (3.13) approaches

$$\alpha_t = \frac{\mathrm{Cov}_t(r_{t+1}, -(\mathrm{E}_{t+1} - \mathrm{E}_t)\sum_{j=0}^{\infty} \rho^j r_{f,t+1+j})}{\sigma_t^2}. \tag{3.16}$$

Now consider the pricing of an inflation-indexed perpetuity or *consol*, that pays one unit of consumption each period forever. Campbell, Lo and

MacKinlay (1997, p. 408), following Shiller (1979), show that a log-linear approximation to the log yield $y_{c,t}$ on a real consol is

$$y_{ct} = \mu_c + (1-\rho_c)\mathrm{E}_t \sum_{j=0}^{\infty} \rho_c^j r_{f,t+1+j}. \tag{3.17}$$

Here μ_c is a constant that captures any risk premium on the consol, and ρ_c is a log-linearization parameter defined as $\rho_c \equiv 1 - \exp\{\mathrm{E}(-p_{c,t})\}$, where $p_{c,t}$ is the log "cum-dividend" price of the consol including its current coupon.[1] Also, the consol return is given by

$$\begin{aligned} r_{c,t+1} &= \frac{1}{1-\rho_c} y_{ct} - \frac{\rho_c}{1-\rho_c} y_{c,t+1} \\ &= r_{f,t+1} + \mu_c - (\mathrm{E}_{t+1} - \mathrm{E}_t) \sum_{j=0}^{\infty} \rho_c^j r_{f,t+1+j}. \end{aligned} \tag{3.18}$$

Thus the consol return has the property that its variance σ_t^2 equals the negative of its covariance with revisions in expected future interest rates. If a consol is the risky asset, $\alpha = 1$ in (3.16), implying that an infinitely risk-averse investor puts all his wealth in an inflation-indexed consol. In this sense the consol, and not the short-term safe asset, is the riskless asset for a long-term investor.

The above argument assumes that $\rho = \rho_c$. These two constants are indeed the same for an individual who is infinitely reluctant to substitute consumption intertemporally ($\psi = 0$). Such an individual consumes the annuity value of wealth, the consumption stream that can be sustained indefinitely by the initial level of wealth. But the annuity value of a real perpetuity is just its dividend of one. Thus for this investor $C/W = 1/P_c$, which implies $\mathrm{E}[c - w] = \mathrm{E}[-p_c]$, and thus, from the definitions of the log-linearization parameters, $\rho = \rho_c$. The infinitely risk-averse investor who is infinitely reluctant to substitute intertemporally holds a real perpetuity that finances a riskless consumption stream over the infinite future.

[1] Campbell, Lo, and MacKinlay work with expected future returns on the perpetuity, but in the current model with constant risk premia these equal the riskless interest rate plus a constant. Also, Campbell, Lo, and MacKinlay give an alternative definition of ρ_c in relation to the "ex-dividend" price of the consol excluding its current coupon. This is more natural in a bond pricing context, but less convenient here because the form of the budget constraint implies that we are measuring wealth inclusive of current consumption, that is, on a "cum-dividend" basis.

3.1.5 Generalizing the solution

The above analysis generalizes straightforwardly to the case where there are multiple risky assets, with or without a short-term riskless asset. As in the previous chapter, we write the variance-covariance matrix of excess risky asset returns as $\mathbf{\Sigma}_t$ and the vector of excess-return variances, the main diagonal of $\mathbf{\Sigma}_t$, as $\boldsymbol{\sigma}_t^2$. We also define a vector $\boldsymbol{\sigma}_{ht}$ that contains the covariances of each risky asset return with declines in expected future interest rates:

$$\boldsymbol{\sigma}_{ht} \equiv \mathrm{Cov}_t(\mathbf{r}_{t+1}, -(\mathrm{E}_{t+1} - \mathrm{E}_t) \sum_{j=0}^{\infty} \rho^j r_{f,t+1+j}). \tag{3.19}$$

The use of the letter h here is intended to evoke Merton's concept of intertemporal hedging demand.

If a short-term riskless asset exists, we have

$$\boldsymbol{\alpha}_t = \frac{1}{\gamma} \mathbf{\Sigma}_t^{-1} (\mathrm{E}_t \mathbf{r}_{t+1} - r_{f,t+1} \boldsymbol{\iota} + \boldsymbol{\sigma}_t^2/2) + \left(1 - \frac{1}{\gamma}\right) \mathbf{\Sigma}_t^{-1} \boldsymbol{\sigma}_{ht}. \tag{3.20}$$

Just as before, the vector of risky asset allocations is a weighted average of a myopic term and a hedging term, where the weights are determined by relative risk tolerance. As risk aversion increases, risk tolerance declines and all weight shifts to the hedging term. If there is no short-term riskless asset, then the solution is augmented by an intercept, just as in the myopic case (2.26). The intercept can be combined with the intertemporal hedging term to write

$$\boldsymbol{\alpha}_t = \frac{1}{\gamma} \mathbf{\Sigma}_t^{-1} (\mathrm{E}_t \mathbf{r}_{t+1} - r_{f,t+1} \boldsymbol{\iota} + \boldsymbol{\sigma}_t^2/2) + \left(1 - \frac{1}{\gamma}\right) \mathbf{\Sigma}_t^{-1} (\boldsymbol{\sigma}_{ht} - \boldsymbol{\sigma}_{0t}), \tag{3.21}$$

where $\boldsymbol{\sigma}_{0t}$ is the vector of covariances of each risky asset's excess return over the benchmark with the benchmark return itself. In practice, as we have noted earlier, the adjustment for benchmark covariances tends to be small when a short-term asset is used as the benchmark.

It is tempting to relate these solutions to the inflation-indexed consol introduced in the previous section. If the loglinearization parameter ρ is fixed and equal to ρ_c, then the vector $\boldsymbol{\sigma}_{ht}$ is equal to a vector $\boldsymbol{\sigma}_{ct}$ containing covariances of each risky asset return with the inflation-indexed consol return. In this case we can give a regression interpretation to the hedging term in (3.20). $\mathbf{\Sigma}_t^{-1} \boldsymbol{\sigma}_{ct}$ is the vector of population regression coefficients from a multiple regression of an inflation-indexed consol return onto the set of risky asset returns. If an inflation-indexed consol is in the set of available risky assets, then this vector will place zero weight on all risky assets except the inflation-indexed consol itself. If an inflation-indexed consol is not available, then the optimal hedging portfolio will be that combination of the risky assets that best approximates the return on an inflation-indexed consol in a regression sense.

The problem with this discussion is that the loglinearization parameter ρ depends on preferences, and will not exactly equal ρ_c unless the investor is infinitely risk-averse and infinitely unwilling to substitute consumption intertemporally ($1/\gamma = \psi = 0$). For all other preferences, $\rho \neq \rho_c$ and thus $\boldsymbol{\sigma}_{ht} \neq \boldsymbol{\sigma}_{ct}$, although the differences are small empirically. In the case where $\psi = 1$, we know that $\rho = \delta$, the time discount factor. If $\psi \neq 1$, we must solve for ρ numerically. We do this using a recursive procedure. We take an initial value of ρ, solve (3.21) for the optimal portfolio, solve for the corresponding optimal consumption-wealth ratio using (3.10), use this consumption-wealth ratio to calculate a new value for ρ, and repeat until convergence. In practice these calculations are extremely rapid and straightforward.

We can also allow for borrowing and short-sales constraints. Unconstrained portfolio allocations are often highly leveraged; but this permits the possibility of bankruptcy in a discrete-time model, and many investors are constrained in their use of leverage. Because the unconstrained optimal portfolio policy is constant over time, we can impose constraints using results in Teplá (1999). Following Cvitanić and Karatzas (1993), Teplá (1999) shows that standard results in static portfolio choice with borrowing and short-sales constraints extend to intertemporal models whose unconstrained optimal portfolio policies are constant over time. The optimal portfolio allocations under borrowing constraints are the unconstrained allocations with a higher short-term interest rate, and the optimal portfolio allocations under short-sales constraints are found by reducing the dimensionality of the asset space until the optimal unconstrained allocations imply no short sales. These and all other results given in this section are explained in detail in the Appendix.

3.2 A Model of the Term Structure of Interest Rates

3.2.1 Specification of the model

Our analysis of portfolio choice with constant risk premia gives a special role to long-term bonds. The riskless asset for a long-term investor is an inflation-indexed consol. If this asset is available it will play some role in the optimal portfolio of any investor who has relative risk aversion greater than one. If an inflation-indexed consol is not available, a conservative investor will hold assets that are good proxies for it.

In order to go further, we need to use an explicit model of the term structure of interest rates to derive quantitative predictions for investors' holdings of Treasury bills, long-term bonds, and equities. Following Campbell and Viceira (2001), we now present such a model. The model is set in discrete time; it has two factors, one real and one nominal, with lognormal distributions and constant variances. This is the simplest term structure model that allows us to distinguish between real and nominal bonds. The real part of the model is a discrete-time version of the well-known Vasicek (1977) continuous-time model.

Because bonds have deterministic payoffs, they can be priced by writing down a time-series model for the stochastic discount factor (SDF) M_{t+1}. The SDF determines the prices of all assets in the economy, but the link is particularly direct with bonds since we do not have to model their payoffs. In a representative-agent framework the SDF can be related to the marginal utility of a representative investor, but here we simply use it as a device to generate a complete set of bond prices. A more detailed explanation of this type of model is given by Campbell, Lo, and MacKinlay (1997, Chapter 11) and Campbell (2000).

We first take logs and work with $m_{t+1} \equiv \log(M_{t+1})$. We break this into its conditional expectation at time t, x_t, and a shock realized at time $t+1$, $v_{m,t+1}$:

$$-m_{t+1} = x_t + v_{m,t+1}. \tag{3.22}$$

We assume that x_t follows an AR(1) process:

$$x_{t+1} = (1 - \phi_x)\,\mu_x + \phi_x x_t + \varepsilon_{x,t+1}. \tag{3.23}$$

We allow the innovations to the log SDF to be correlated with innovations to its conditional expectation:

$$v_{m,t+1} = \beta_{mx}\varepsilon_{x,t+1} + \varepsilon_{m,t+1}. \tag{3.24}$$

The term structure of real interest rates

The economic meaning of these assumptions can best be appreciated by working out their implications for the term structure of real interest rates. There is a direct link between the stochastic discount factor and the log return, or equivalently the log yield, on a one-period indexed bond: $r_{1,t+1} = -\log \mathrm{E}_t[M_{t+1}] = \mathrm{E}_t[-m_{t+1}] - \frac{1}{2}\mathrm{Var}_t[m_{t+1}]$ since M_{t+1} is lognormal. Substituting in from (3.22)–(3.24), we have

$$r_{1,t+1} = x_t - \frac{1}{2}\left(\beta_{mx}^2\sigma_x^2 + \sigma_m^2\right), \tag{3.25}$$

where $\sigma_x^2 \equiv \mathrm{Var}_t(\varepsilon_{x,t+1})$ and $\sigma_m^2 \equiv \mathrm{Var}_t(\varepsilon_{m,t+1})$. The short-term real interest rate equals the state variable x_t adjusted by a constant, so it follows an AR(1) process with persistence ϕ.

Longer-term inflation-indexed bonds can be priced recursively. A two-period bond today will become a one-period bond tomorrow, when its price will be described by (3.25). Thus one can solve for its log price, or equivalently its log yield, and then the log prices and yields of all longer-maturity bonds. The solution for the log yield on an n-period indexed zero-coupon bond, y_{nt}, times bond maturity n, which equals minus the log price of the bond, p_{nt}, is given by

$$ny_{nt} = -p_{nt} = A_n + B_n x_t, \tag{3.26}$$

where A_n and B_n are functions of bond maturity n but not of time t.

A recursive expression for the coefficient A_n is given in the Appendix. More important for our present discussion, the coefficient B_n is given by

$$B_n = 1 + \phi_x B_{n-1} = \frac{1-\phi_x^n}{1-\phi_x}. \tag{3.27}$$

To understand this equation, note that the expectations hypothesis of the term structure holds in this model. Thus the log bond yield is a constant plus an average of expected future real interest rates over the next n periods, and n times the log bond yield is a constant plus a sum of expected future real interest rates. Since the real interest rate follows an AR(1) process with persistence coefficient ϕ_x, the expected real interest rate k periods ahead is a constant plus ϕ_x^k times the real interest rate today. Summing up over n periods gives (3.27).

The one-period log return on an n-period indexed zero-coupon bond is just the change in its price, $(p_{n-1,t+1} - p_{n,t})$. Combining this expression with (3.26) and (3.27), the excess return over the one-period log interest rate is

$$r_{n,t+1} - r_{1,t+1} = -\frac{1}{2}B_{n-1}^2\sigma_x^2 - \beta_{mx}B_{n-1}\sigma_x^2 - B_{n-1}\varepsilon_{x,t+1}, \tag{3.28}$$

so the n-period bond is sensitive only to real-interest-rate shocks $\varepsilon_{x,t+1}$, with a sensitivity B_{n-1}. Because the real part of our model has only a single factor, yields and prices of inflation-indexed bonds of all maturities are driven only by the short-term real interest rate and thus are perfectly correlated with one another. The variance of the n-period bond return is $\sigma_{nt}^2 = B_{n-1}^2 \sigma_x^2$, and the risk premium is

$$\mathrm{E}_t \left[r_{n,t+1} - r_{1,t+1} \right] + \frac{\sigma_{nt}^2}{2} = -\beta_{mx} B_{n-1} \sigma_x^2. \tag{3.29}$$

The risk premium on the long-term bond, or term premium, is determined by the conditional covariance of the excess bond return with the log SDF. In our homoskedastic model the conditional covariance is constant through time but dependent on the bond maturity; thus the term premium is constant as postulated by the expectations hypothesis of the term structure. Since $B_{n-1} > 0$, the term premium has the opposite sign to β_{mx}. With a positive β_{mx}, long-term indexed bonds pay off when the SDF or, equivalently, the marginal utility of consumption for a representative investor is high, that is, when wealth is most desirable. In equilibrium, these bonds have a negative term premium and the real yield curve is on average downward-sloping. With a negative β_{mx}, on the other hand, long-term indexed bonds pay off when the marginal utility of consumption for a representative investor is low, and so in equilibrium they have a positive term premium. In this case the real yield curve is on average upward-sloping.

Equation (3.29) implies that the Sharpe ratio for indexed bonds is $-\beta_{mx}\sigma_x$, which is independent of bond maturity. The invariance of the Sharpe ratio to bond maturity follows from the single-factor structure of the real sector of the model. The ratio of the risk premium to the variance of the excess return, which determines a myopic investor's allocation to long-term bonds, is $-\beta_{mx}/B_{n-1}$. This does depend on bond maturity but not on the volatility of the real interest rate.

The optimal portfolio of inflation-indexed bonds

If only inflation-indexed bonds are available to the investor, the real sector of the model is all we need to derive an explicit solution to the portfolio problem. Because all long-term inflation-indexed bonds are perfectly correlated, we can consider without loss of generality the choice between two assets: a single-period inflation-indexed bond and an n-period inflation-indexed bond. Using (3.28) to calculate the terms in (3.13), we find that the optimal portfolio weight on the n-period bond is

$$\alpha_n = \frac{1}{\gamma} \left(\frac{-\beta_{mx}}{B_{n-1}} \right) + \left(1 - \frac{1}{\gamma} \right) \left(\frac{1}{B_{n-1}} \right) \left(\frac{\rho}{1 - \rho \phi_x} \right). \tag{3.30}$$

This solution has several interesting properties. First, the variance of real interest rate shocks, σ_x^2, does not directly affect the portfolio weight α_n because, given the parameterization of our model, it moves the variance of returns and the risk premium in proportion to one another. Interest-rate variance can only affect the solution indirectly, through the loglinearization parameter ρ. Empirically we find that ρ changes very little when the parameters of the model change, and so the indirect effect through ρ is quantitatively negligible.

Second, the interest-rate sensitivity of the optimal portfolio is given by $\alpha_n B_{n-1}$, and this does not depend on the bond maturity n. Two assets are enough to complete the market with respect to real-interest-rate risk, and thus the investor can use any two inflation-indexed bonds to construct a portfolio with the optimal level of interest-rate sensitivity. If only short-maturity bonds with a low sensitivity are available, the investor can compensate by holding more of them or even leveraging his position; if only long-maturity bonds with a high sensitivity are available, the investor can compensate by holding fewer of them.

Third, the intertemporal hedging demand, the second term in (3.30), is increasing in the persistence of interest-rate shocks ϕ_x. This is because, as shown in the general solution (3.13), hedging demand is determined by the covariance of the risky asset return with the discounted value of all future interest rates. A major theme of the empirical work in this book is that persistent shocks to investment opportunities are much more important for portfolio choice than are transitory shocks to investment opportunities.

Fourth, as risk aversion increases and the elasticity of intertemporal substitution declines, the limit of (3.30) is a portfolio that is equivalent to an inflation-indexed consol. In this model an inflation-indexed consol has interest-rate sensitivity $-\rho_c/(1-\rho_c\phi_x)$, and $\rho = \rho_c$ when $1/\gamma = \psi = 0$. This is a special case of the general point made in section 3.1.4.

Equation (3.30) is appealingly simple, but it lacks realism because it does not allow investors to hold nominal bonds or equities. In order to go further, we need to augment the model to include such assets.

The term structure of nominal interest rates

In order to price nominal bonds, we must model the process driving inflation. We assume that this process has the same form that we have already assumed for the SDF. That is, realized log inflation π_{t+1} equals expected log inflation z_t plus an inflation shock, and expected inflation follows an

AR(1) process:

$$\pi_{t+1} = z_t + v_{\pi,t+1}, \tag{3.31}$$
$$z_{t+1} = (1-\phi_z)\,\mu_z + \phi_z z_t + v_{z,t+1}. \tag{3.32}$$

We assume that the shocks to realized and expected inflation, $v_{\pi,t+1}$ and $v_{z,t+1}$, can be correlated with each other and with the real shocks to the model:

$$v_{z,t+1} = \beta_{zx}\varepsilon_{x,t+1} + \beta_{zm}\varepsilon_{m,t+1} + \varepsilon_{z,t+1}, \tag{3.33}$$
$$v_{\pi,t+1} = \beta_{\pi x}\varepsilon_{x,t+1} + \beta_{\pi m}\varepsilon_{m,t+1} + \beta_{\pi z}\varepsilon_{z,t+1} + \varepsilon_{\pi,t+1}, \tag{3.34}$$

The model is driven by four normally distributed, white noise shocks $\varepsilon_{m,t+1}$, $\varepsilon_{\pi,t+1}$, $\varepsilon_{x,t+1}$, and $\varepsilon_{z,t+1}$ that determine the innovations to the log SDF, the log inflation rate, and their conditional means. These shocks are cross-sectionally uncorrelated, with variances σ_m^2, σ_π^2, σ_x^2, and σ_z^2. It is important to note that z_{t+1}, the expected inflation rate, is affected by both a pure expected-inflation shock $\varepsilon_{z,t+1}$ and the shocks to the expected and unexpected log SDF $\varepsilon_{x,t+1}$ and $\varepsilon_{m,t+1}$. That is, innovations to expected inflation can be correlated with innovations in the log SDF, and hence with innovations in the short-term real interest rate. These correlations mean that nominal interest rates need not move one-for-one with expected inflation—that is, the Fisher hypothesis need not hold—and nominal bond prices can include an inflation risk premium as well as a real term premium.

We have written the model with a self-contained real sector (3.22)–(3.24) and a nominal sector (3.31)–(3.34) that is affected by shocks to the real sector. But this is merely a matter of notational convenience. Our model is a reduced form rather than a structural model, so it captures correlations among shocks to real and nominal interest rates but does not have anything to say about the true underlying sources of these shocks.

The pricing of nominal bonds follows the same steps as the pricing of indexed bonds. The log price of an n-period nominal zero-coupon bond, $p_{n,t}^{\$}$, is a linear combination of x_t and z_t whose coefficients are time-invariant, though they vary with the maturity of the bond:

$$-p_{n,t}^{\$} = A_n^{\$} + B_{1,n}^{\$} x_t + B_{2,n}^{\$} z_t. \tag{3.35}$$

The Appendix gives expressions for the coefficients $A_n^{\$}$, $B_{1,n}^{\$}$ and $B_{2,n}^{\$}$.

Since nominal bond prices are driven by shocks to both real interest rates and inflation, they have a two-factor structure rather than the single-factor structure of indexed bond prices. Inflation affects the excess return on an n-period nominal bond over the one-period nominal interest rate, so risk premia in the nominal term structure include compensation for inflation risk. Like all other risk premia in the model, however, the risk premia on nominal bonds are constant over time; thus the expectations hypothesis of the term structure holds for nominal as well as for real bonds.

Pricing equities

Even though our focus in this chapter is on long-term bonds, a full evaluation of bond demand requires that we include equities in our model as an attractive alternative long-term investment. We do this as simply as possible, assuming that the unexpected log return on equities is affected by shocks to both the expected and unexpected log SDF:

$$r_{e,t+1} - \mathrm{E}_t\, r_{e,t+1} = \beta_{ex}\varepsilon_{x,t+1} + \beta_{em}\varepsilon_{m,t+1}. \tag{3.36}$$

Campbell (1999) shows that this decomposition of the unexpected log equity return into a linear combination of the shocks to the expected and unexpected log SDF is consistent with a representative-agent endowment model where expected aggregate consumption growth follows an AR(1). From the fundamental pricing equation $1 = \mathrm{E}_t[M_{t+1}R_{t+1}]$ and the lognormal structure of the model it is easy to show that the risk premium on equities, over a one-period riskless return $r_{1,t+1}$, is given by

$$\mathrm{E}_t[r_{e,t+1} - r_{1,t+1}] + \frac{\sigma_{et}^2}{2} = \beta_{mx}\beta_{ex}\sigma_x^2 + \beta_{em}\sigma_m^2. \tag{3.37}$$

Like all other covariances in the model, this is constant over time so the equity premium is a constant.

Once nominal bonds and equities are included in the model, the algebraic portfolio solutions become sufficiently complicated that it is no longer helpful to write them out explicitly. Instead, we estimate the model on historical data and present numerical solutions.

3.2.2 The term structure of interest rates in the US

Data and estimation method

We estimate the two-factor term structure model using data on US nominal interest rates, equities and inflation. We use nominal zero-coupon yields at maturities 3 months, 1 year, 3 years, and 10 years from McCulloch and Kwon (1993), updated by Gong and Remolona (1996a,b). We take data on equities from the Indices files on the CRSP tapes, using the value-weighted return, including dividends, on the NYSE, AMEX and NASDAQ markets. For inflation, we use a Consumer Price Index that retrospectively incorporates the rental-equivalence methodology, thereby avoiding any direct effect of nominal interest rates on measured inflation. The Appendix shows that estimation results are extremely similar if we instead use the personal consumption expenditure (PCE) deflator to measure inflation. Although the raw data are available monthly, we construct a quarterly data set in order to reduce the influence of high-frequency noise in inflation and short-term movements in interest rates. We begin our sample in 1952, just after the Fed-Treasury Accord that dramatically altered the time-series behavior of nominal interest rates. Our data end in 1996.

To avoid the implication of the model that bond returns are driven by only two common factors, so that all bond returns can be perfectly explained by any two bond returns, we assume that bond yields are measured with error. The errors in yields are normally distributed, serially uncorrelated, and uncorrelated across bonds. Then the term structure model becomes a classic state-space model in which unobserved state variables x_t and z_t follow a linear process with normal innovations and we observe linear combinations of them with normal errors. The model can be estimated by maximum likelihood using a Kalman filter to construct the likelihood function.

In Table 3.1 we report parameter estimates for the period 1952-96 and the period 1983-96. Interest rates were unusually high and volatile in the 1979-82 period, during which the Federal Reserve Board under Paul Volcker was attempting to reestablish the credibility of anti-inflationary monetary policy and was experimenting with monetarist operating procedures. Many authors have argued that real interest rates and inflation have behaved differently in the monetary policy regime established since 1982 by Federal Reserve chairmen Volcker and Alan Greenspan (see for example Clarida, Gali and Gertler 1998). Accordingly we report separate estimates for the period starting in 1983 in addition to the full sample period.

The parameter values in Table 3.1 are restricted maximum likelihood estimates of the model. Unrestricted maximum likelihood estimates fit the data well in the 1952-96 sample period, but they deliver implausibly low means for short-term nominal and real interest rates in the 1983-96 sample period. (The model does not necessarily fit the sample means because the same parameters are used to fit both time-series and cross-sectional behavior; thus the model can trade off better fit elsewhere for worse fit of mean short-term interest rates.) Accordingly we require that the model exactly fit the sample means of nominal interest rates and inflation. This restriction

Table 3.1: Term Structure Model Estimation

	1952.I - 1996.III		1983.I - 1996.III	
Parameter	est.	s.e.	est.	s.e.
μ_x	0.0573	0.0298	0.0194	0.0693
μ_z	0.0094		0.0087	
ϕ_x	0.8688	0.0057	0.9862	0.0042
ϕ_z	0.9992	0.0012	0.8599	0.0216
β_{mx}	-74.9797	41.6949	-28.6919	114.0025
β_{zx}	0.0752	0.0516	-0.4114	0.1886
β_{zm}	-0.0012	0.0006	0.0008	0.0024
$\beta_{\pi x}$	0.5198	0.3050	-0.0267	0.9790
$\beta_{\pi m}$	-0.0088	0.0034	0.0008	0.0193
$\beta_{\pi z}$	1.4320	0.2940	-1.5412	1.5047
β_{ex}	-3.4957	3.4123	-9.3629	6.3014
β_{em}	0.3013	0.0979	0.5089	1.3528
σ_x	0.0025	0.0001	0.0027	0.0006
σ_m	0.2694	0.0927	0.1351	0.3579
σ_z	0.0013	0.0001	0.0016	0.0002
σ_π	0.0071	0.0004	0.0072	0.0018
log-lik.	26.3327		26.8222	
no. obs.	179		55	
$E[r_{1,t+1}]$	1.39%		2.93%	
$E[r^{\$}_{1,t+1}]$	5.50%		6.40%	
$\sigma(r_{1,t+1})$	1.01%		3.25%	
$\sigma(r^{\$}_{1,t+1})$	6.70%		3.09%	
$E[\pi_{t+1}]$	3.77%		3.49%	
$\sigma_t(\pi_{t+1})$	1.57%		1.52%	

hardly reduces the likelihood at all in 1952-96, and even in 1983-96 it cannot be rejected at conventional significance levels.

Parameter estimates

The first two columns of Table 3.1 report parameters and asymptotic standard errors for the period 1952-96. All parameters are in natural units, so they are on a quarterly basis. We estimate a moderately persistent process for the real interest rate; the persistence coefficient ϕ_x is 0.87, implying a half-life for shocks to real interest rates of about 5 quarters. The expected inflation process is much more persistent, with a coefficient ϕ_z of 0.9992 that implies a half-life for expected inflation shocks of over two centuries! Of course, the model also allows for transitory noise in realized inflation.

The bottom of Table 3.1 reports the implications of the estimated parameters for the means and standard deviations of real interest rates, nominal interest rates, and inflation, measured in percent per year. The implied mean log yield on an indexed three-month bill is 1.39 percent for the 1952-96 sample period. Taken together with the mean log yield on a nominal three-month bill of 5.50 percent and the mean log inflation rate of 3.77 percent (both restricted to equal the sample means over this period), and adjusting for Jensen's Inequality using one-half the conditional variance of log inflation, the implied inflation risk premium in a three-month nominal Treasury bill is 35 basis points. This fairly substantial risk premium is explained by the significant positive coefficient $\beta_{\pi x}$ and the significant negative coefficient $\beta_{\pi m}$ in Table 3.1.[7]

Risk premia on long-term indexed bonds, relative to a three-month indexed bill, are determined by the parameter β_{mx}. This is negative and significant, implying positive risk premia on long-term indexed bonds and an upward sloping term structure of real interest rates. Risk premia on nominal bonds, relative to indexed bonds, are determined by the inflation-risk parameters β_{zx} and β_{zm}. The former is positive but statistically insignificant, while the latter is negative and significant. Both point estimates imply positive inflation risk premia on nominal bonds relative to indexed bonds.

The parameters in Table 3.1 can also be used to calculate the volatility of the log stochastic discount factor. From (3.22)–(3.24), the variance of m_{t+1} is $\sigma_x^2/(1-\phi_x^2)+\beta_{mx}^2\sigma_x^2+\sigma_m^2$. The estimates in Table 3.1 imply a large quarterly standard deviation of 0.33, consistent with the literature on volatility bounds for the stochastic discount factor (Hansen and Jagannathan 1991, Cochrane and Hansen 1992). When financial markets are complete, the discounted marginal utility growth of each investor must be equal to the stochastic discount factor. Therefore the consumption and portfolio solutions we report later in the paper for the complete-markets case imply highly volatile marginal utilities, due either to volatile consumption or high risk aversion. This is a manifestation of the equity premium puzzle of Mehra and Prescott (1985) in our microeconomic model with exogenous asset returns and endogenous consumption.

Implications for the term structure

Table 3.2 explores the term-structure implications of our estimates in greater detail. The table compares implied and sample moments of term structure variables, measured in percent per year. It also reports standard errors for the implied moments, calculated using the delta method. Panel A of Table 3.2 reports sample moments for returns and yields on nominal bonds, together with the moments implied by our estimated model; panel B shows comparable implied moments for indexed bonds, and panel C reports sample and implied moments for equities. Row 1 of the table gives Jensen's-Inequality-corrected average excess returns on n-period nominal bonds over 1-period nominal bonds, while row 2 gives the standard deviations of these excess returns. Row 3 reports annualized Sharpe ratios for nominal bonds, the ratio of row 1 to row 2. Row 4 reports mean nominal yield spreads and row 5 reports the standard deviations of nominal yield spreads. Rows 6 through 10 repeat these moments for indexed bonds. Note that the reported risk premia and Sharpe ratios for nominal and indexed bonds are not directly comparable because they are measured relative to different short-term assets, nominal and indexed respectively.

A comparison of the model implications in rows 1 and 6 shows that 10-year nominal bonds have a risk premium over three-month nominal bills of 1.97 percent per year, while 10-year indexed bonds have a risk premium over three-month indexed bills of 1.35 percent per year. These numbers, together with the 35-basis-point risk premium on three-month nominal bills over three-month indexed bills, imply a 10-year inflation risk premium (the risk premium on 10-year nominal bonds over 10-year indexed bonds) slightly above 1.1 percent. This estimate is consistent with the rough calculations in Campbell and Shiller (1996).

Rows 2 and 7 show that nominal bonds are much more volatile than indexed bonds; the difference in volatility increases with maturity, so that 10-year nominal bonds have a standard deviation three times greater than 10-

Table 3.2: Sample and Implied Moments of the Term Structure

	Moment		1952.I - 1996.III		1983.I - 1996.III	
			1 yr.	10 yr.	1 yr.	10 yr.
		A: Nominal Term Structure				
(1)	$E[r^{\$}_{n,t+1} - r^{\$}_{1,t+1}] + \sigma^{2\$}_n/2$	sample	0.397	0.915	0.706	5.675
		implied	0.559	1.967	0.155	2.278
(2)	$\sigma(r^{\$}_{n,t+1} - r^{\$}_{1,t+1})$	sample	1.615	11.365	1.135	12.612
		implied	1.634	11.566	1.312	14.896
(3)	$SR^{\$} = (1)/(2)$	sample	0.246	0.080	0.622	0.450
		implied	0.342	0.170	0.118	0.153
(4)	$E[y^{\$}_{n,t+1} - y^{\$}_{1,t+1}]$	sample	0.440	1.185	0.527	2.067
		implied	0.294	1.174	0.071	0.766
(5)	$\sigma(y^{\$}_{n,t+1} - y^{\$}_{1,t+1})$	sample	0.222	0.613	0.177	0.545
		implied	0.182	0.826	0.140	0.803
		B: Real Term Structure				
(6)	$E[r_{n,t+1} - r_{1,t+1}] + \sigma^2_n/2$	implied	0.490	1.345	0.245	2.513
(7)	$\sigma(r_{n,t+1} - r_{1,t+1})$	implied	1.309	3.788	1.590	16.295
(8)	$SR = (6)/(7)$	implied	0.374	0.374	0.154	0.154
(9)	$E[y_{n,t+1} - y_{1,t+1}]$	implied	0.253	1.100	0.118	0.858
(10)	$\sigma(y_{n,t+1} - y_{1,t+1})$	implied	0.182	0.816	0.067	0.738
		C: Equities				
(11)	$E[r_{e,t+1} - (r^{\$}_{1,t+1} - \pi_{t+1})]$	sample		6.910		8.738
	$+\sigma^2(r_{e,t+1} - (r^{\$}_{1,t+1} - \pi_{t+1}))/2$	implied		8.988		4.527
(12)	$\sigma(r_{e,t+1} - (r^{\$}_{1,t+1} - \pi_{t+1}))$	sample		15.917		14.646
		implied		15.896		14.748
(13)	$SR = (11)/(12)$	sample		0.434		0.597
		implied		0.565		0.307

year indexed bonds. This difference in volatility makes the Sharpe ratio for indexed bonds in row 8 considerably higher than the Sharpe ratio for nominal bonds in row 3. Since indexed bond returns are generated by a single-factor model, the Sharpe ratio for indexed bonds is independent of maturity at 0.37. The Sharpe ratio for nominal bonds declines with maturity; short-term nominal bonds have a ratio close to that for indexed bonds, but the Sharpe ratio for 10-year nominal bonds is only 0.17. These numbers imply that in our portfolio analysis, investors with low risk aversion will have a strong myopic demand for indexed bonds.

Table 3.2 can also be used to evaluate the empirical fit of the model. A comparison of the model's implied moments with the sample moments for nominal bonds shows that the model fits the volatility of excess nominal bond returns and changes in yields extremely well. The model somewhat overstates the average excess nominal bond return and the nominal Sharpe ratio, but this can be attributed in part to the upward drift in interest rates over the 1952-96 sample period which biases downward the sample means. The standard errors for implied volatilities are small, while the standard errors of implied mean excess returns are large. This reflects the well-known result that it is much harder to obtain precise estimates of first moments than of second moments.

Another way to judge the fit of the model is to ask how much of the variability of bond yields, or bond returns, is accounted for by the structural parameters as opposed to the white-noise measurement errors we have allowed in each bond yield. The estimated variances of measurement errors (not reported in Table 3.1) are zero for 1-year and 10-year bonds and are extremely small for 3-month bills and 3-year bonds. Measurement errors are estimated to account for less than 0.5 percent of the variance of 3-month and 3-year bond yields and less than 5 percent of the variance of 3-year bond returns. This reflects the fact that bond yields are highly persistent at all maturities, so the model fits them primarily with persistent structural processes rather than white-noise measurement errors.

Finally, we can use the model to describe the history of the nominal US term structure and its components. We do this in Figure 3.1, which shows 3-month nominal bill yields, real yields and expected inflation in the top panel, and the equivalent 10-year series in the bottom panel. The model attributes the low-frequency variation of nominal interest rates, particularly the runup in nominal rates from the 1960's through the early 1980's and the slow decline thereafter, to changing expected inflation. Much of the higher-frequency variation in interest rates is attributed to the real interest rate, particularly after the end of the 1960's.[2] Because real-interest-rate variation is less persistent than expected-inflation variation, the latter is the main determinant of the 10-year nominal yield. However the 10-year real yield does show some residual variation. Overall the model appears to provide a good description of the nominal US term structure considering its parsimony and the fact that we have forced it to fit both time-series and cross-sectional features of the data.

Rows 11, 12, and 13 of Table 3.2 report summary statistics for equities: the annualized Jensen's-Inequality-corrected average excess returns on equities relative to nominal bills, the standard deviation of these excess returns, and their Sharpe ratio. The model fits the standard deviation of equities extremely well but overpredicts the equity premium and the Sharpe ratio for equities. The implied Sharpe ratio of 0.57 implies that investors with low risk aversion will have an extremely large myopic demand for equities; this is again a manifestation of the equity premium puzzle.

The right hand sides of Tables 1 and 2 repeat these estimates for the Volcker-Greenspan period 1983–96. Many of the parameter estimates are quite similar; however we find that in this period real interest rates are much more persistent, with $\phi_x = 0.986$ and an implied half-life for real interest rate shocks of about 12 years. The expected inflation process now mean-reverts much more rapidly, with $\phi_z = 0.860$ implying a half-life for expected inflation shocks of about 5 quarters. These results are consistent with the patterns illustrated in Figure 3.1, and with the notion that since the early 1980's the Federal Reserve has more aggressively controlled inflation at the cost of greater long-term variation in the real interest rate (Clarida, Gali, and Gertler 1998). The increase in real-interest-rate persistence increases the risk premia on indexed and nominal bonds, but it also greatly increases the volatility of indexed bond returns so the Sharpe ratio for indexed bonds is lower at 0.15. In the remainder of the chapter we present portfolio choice results based on our full-sample estimates for the period 1952–96, but we also discuss results for the 1983–96 period where they are importantly different.

[2]Fama (1975) famously argued that the real interest rate is constant and that all variation in nominal interest rates is due to expected inflation. This was a reasonable view of his data, which ended in 1971, but certainly does not describe more recent experience. Perhaps Fama fell foul of "Murphy's Law of Empirical Economics", that any strong characteristic of historical data will alter immediately after it has been identified by empirical economists!

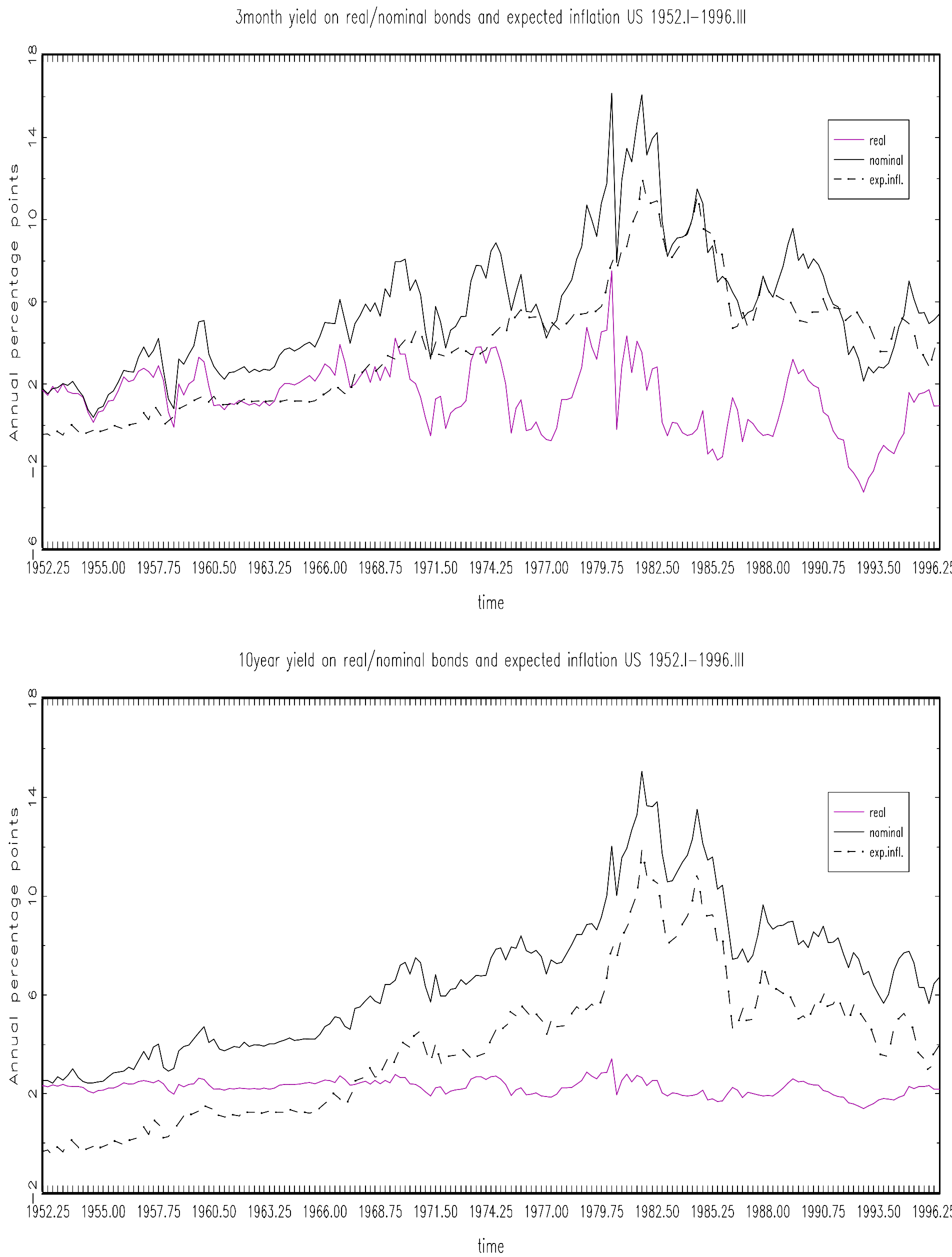

Figure 3.1: Fitted real and nominal yields and inflation

Table 3.3: Optimal Allocations to Equities and Long-Term Bonds

	Unconstrained		Constrained		Unconstrained		Constrained	
R.R.A.	Equity	Indexed	Equity	Indexed	Equity	Nominal	Equity	Nominal
				(A) Sample Period: 1952 - 1996				
0.75	443	1082	100	0	470	25	100	0
1	332	835	100	0	352	21	100	0
2	166	464	100	0	175	15	100	0
5	66	242	60	40	69	12	69	12
10	33	168	30	70	33	11	33	11
5000	0	94	0	94	-2	10	0	10
				(B) Sample Period: 1983 - 1996				
0.75	262	-1	100	0	259	1	100	0
1	196	21	94	6	195	24	96	4
2	98	54	53	47	99	58	52	48
5	39	74	28	72	41	78	25	75
10	20	81	19	81	22	85	16	84
5000	0	88	0	88	3	92	3	92

3.2.3 Implications for portfolio choice

Given our estimates of the term structure model, it is straightforward to calculate all the terms in the intertemporal portfolio solution (3.21). Table 3.3 reports optimal demands for equities and for 3-month and 10-year indexed or nominal bonds by investors who are unconstrained or subject to borrowing and short-sales constraints. For simplicity we assume either that short- and long-term bonds are all indexed, or that they are all nominal; we do not allow investors to hold equities, indexed bonds, and nominal bonds simultaneously. Panel A reports results for the 1952–96 sample, which we consider first.

In a world with full indexation, the unconstrained demand for both long-term indexed bonds and equities is positive and often above 100 percent, implying that the investor optimally borrows to finance purchases of equities and indexed bonds. The portfolio share of indexed bonds exceeds that of equities, despite the higher Sharpe ratio of equities, because indexed bonds are much less risky than equities.[3] As the coefficient of relative risk aversion increases, the demands for both long-term indexed bonds and equities fall, but the share of equities falls faster. In the limit the infinitely risk-averse investor holds a portfolio equivalent to an indexed perpetuity as we have already discussed. When there are borrowing and short-sale constraints, investors with low risk aversion invest fully in equities as a way to maximize their risk and expected return without using leverage, while more risk-averse investors hold both indexed bonds and equities. Cash plays only a minor role and only in the portfolios of the most risk-averse investors, who are almost fully invested in indexed bonds.

These findings are related to the "asset allocation puzzle" of Canner, Mankiw, and Weil (1997) discussed in Chapter 1. Investment advisers often suggest that more conservative investors should have a higher ratio of long-term bonds to stocks in their portfolios. Canner, Mankiw, and Weil document this feature of conventional investment advice and point out that it is inconsistent with the mutual fund theorem of static portfolio analysis, according to which risk aversion should affect only the ratio of cash to risky assets and not the relative weights on different risky assets.

Our analysis shows that static portfolio analysis can be seriously misleading when investment opportunities are time-varying and investors have long time horizons. The portfolio allocations to equities and indexed bonds in Panel A of Table 3.3 are strikingly consistent with conventional investment advice. Aggressive long-term investors should hold stocks, while conservative ones should hold long-term bonds and small amounts of cash. The explanation is that long-term bonds, and not cash, are the riskless asset for long-term investors.

A weakness in this resolution of the asset allocation puzzle is that it assumes that long-term bonds are indexed, or equivalently, that there is no inflation uncertainty. Panel A of Table 3.3 shows that nominal bonds play a much smaller role in optimal portfolios. In a world with no indexation,

[3]Recall that optimal myopic demand for a single risky investment, or for a risky investment that is independent of other risky investments, is proportional to mean excess return divided by variance. Equivalently, it is proportional to Sharpe ratio divided by standard deviation. Although equities have a higher Sharpe ratio than indexed bonds, their standard deviation is much higher so the optimal equity share is lower.

unconstrained investors with low risk aversion hold modest nominal bond positions, but constrained investors hold only equities. As risk aversion increases, investors move into cash rather than long-term nominal bonds. Figure 3.2 illustrates this point. The top panel of the figure plots constrained allocations to equities, a ten-year nominal bond and a three-month nominal bill, while the bottom panel plots constrained myopic allocations. The horizontal axis measures relative risk tolerance $(1/\gamma)$ rather than relative risk aversion, because both total and myopic allocations are linear in risk tolerance when portfolio constraints are not binding. Infinitely conservative investors with $1/\gamma = 0$ are plotted at the right edge of the figure. As in the tables we set $\psi = 1$, but the choice of ψ has very little effect on the results.

Risk-tolerant investors at the left of Figure 3.2 are fully invested in equities. Highly risk-averse investors hold most of their portfolios in cash, although they also hold some ten-year nominal bonds. The bottom panel shows that long-term bonds are held purely for hedging purposes. The myopic demand for long-term bonds is extremely close to zero at all levels of risk aversion.

The portfolio allocations to nominal bonds in Panel A of Table 3.3 and Figure 3.2 do not correspond well with conventional investment advice. In order to rationalize the conventional wisdom about long-term nominal bonds, one must assume that future interest rates will be generated by a different process than the one estimated in 1952–96, a process with less uncertainty about future inflation. Interestingly, we have estimated just such a process over the Volcker-Greenspan sample period 1983-96. Panel B of Table 3.3 repeats Panel A using our 1983-96 estimates and finds that even when only nominal bonds are available, aggressive long-term investors should hold stocks, while conservative ones should hold primarily long-term nominal bonds along with small quantities of stocks.[4]

Figure 3.3, whose structure is identical to Figure 3.2, emphasizes this result. Panel A shows that almost all investors should be fully invested in equities and long-term nominal bonds when they face a term structure like the one estimated for the Volcker-Greenspan era. Only extremely risk-averse

[4] During the 1983-96 period the interest-rate sensitivity of a 10-year indexed zero-coupon bond is considerably less than that of an indexed perpetuity. Therefore an infinitely risk-averse investor would like to hold a leveraged position in 10-year indexed zeros, which was not the case in our 1952-96 model. To maintain comparability with that model, in Panel B of Table 3.3 and Figure 3.3 we replace the 10-year zero-coupon bond with a 20-year zero-coupon bond. This ensures that the optimal indexed portfolio for an infinitely risk-averse investor is available even when borrowing and short-sales constraints are imposed.

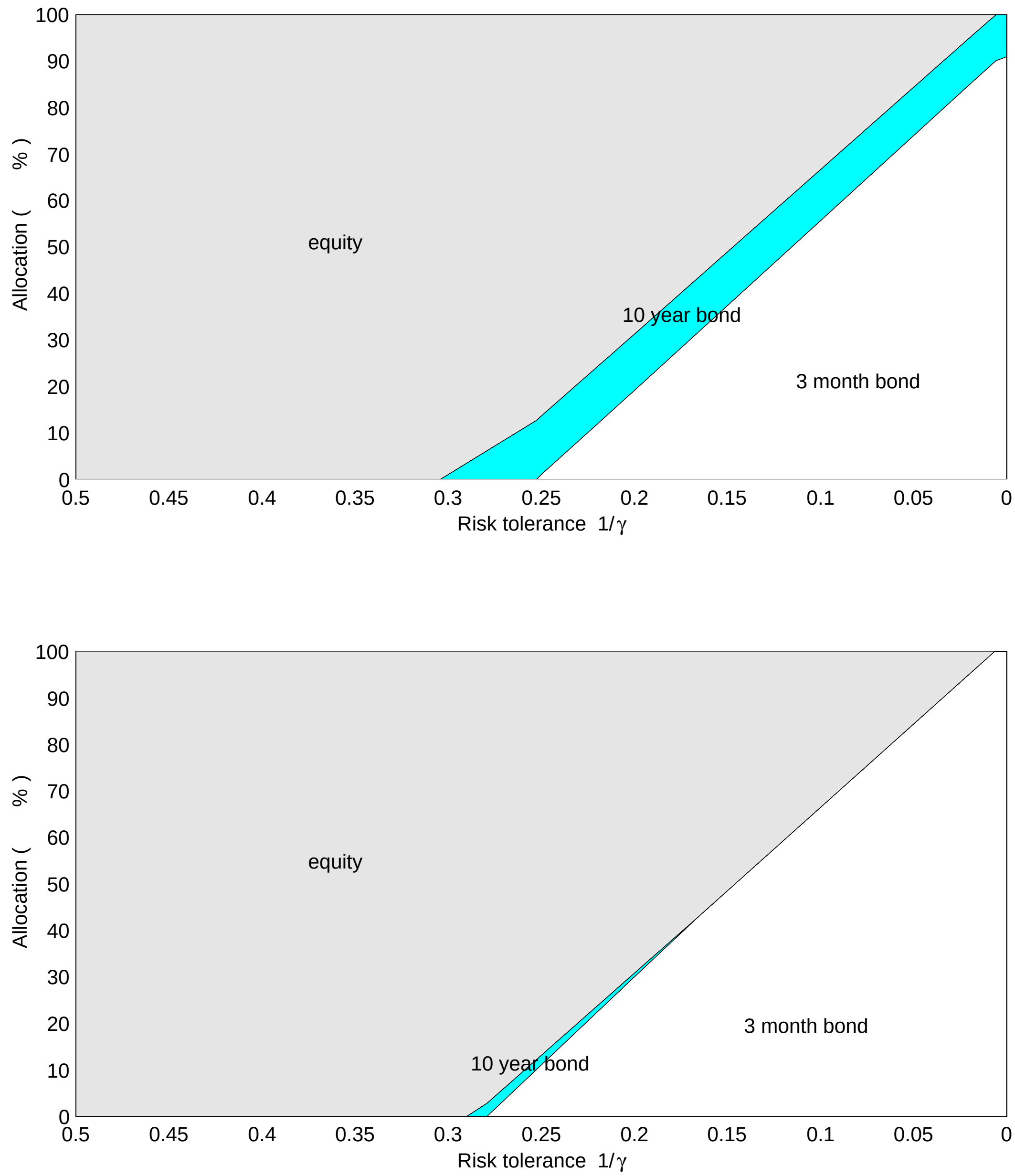

Figure 3.2: Optimal portfolios for the US 1952-96

investors should hold some cash in their portfolios. Panel B shows that intertemporal hedging motives account for most of this demand for long-term nominal bonds. If investors behaved myopically and ignored the hedging properties of long-term bonds, their portfolios would contain mostly equities and cash. The top panel of Figure 3.3 also shows that the ratio of nominal bonds to equities in the optimal portfolio increases with risk aversion, just as recommended by conventional investment advice. If investors behaved myopically, this ratio would be constant when portfolio constraints are not binding, as shown in the bottom panel.

Although our 1983-96 model replicates important features of conventional investment advice, it still falls short in one respect. The optimal portfolios in Figure 3.3 contain very little cash relative to the recommended portfolios reported by Canner, Mankiw, and Weil (1997). We do not attempt to match those portfolios more accurately, but suspect that it can be done either by using a term-structure process intermediate between the two processes we have estimated, or by modelling liquidity motives for holding cash.

3.3 Conclusion: Bonds, James, Bonds

If one uses conventional mean-variance analysis, it is hard to explain why any investors hold large positions in bonds. Mean-variance analysis treats cash as the riskless asset, and treats bonds merely as another risky asset like stocks. Bonds are valued only for their potential contribution to the short-run excess return, relative to risk, of a diversified risky portfolio. This view tends to relegate bonds to a minor supporting role in the recommended portfolio, since excess bond returns have historically been fairly low and bond returns have been highly variable in the short run. Over the period 1952–96 reported in Table 3.2, for example, the average excess return on 10-year US Treasury zero-coupon bonds over 3-month Treasury bills was less than 1% while the standard deviation of this return was over 11%. Accordingly the annualized Sharpe ratio for bonds, the ratio of their average excess return to their standard deviation, was less than 0.1. Over the same period the US equity market had an average excess return of almost 7% and a standard deviation of 16%, implying a Sharpe ratio above 0.4. The comparison looks somewhat more favorable for bonds in the shorter 1983–96 sample also shown in Table 3.2, but it is even less favorable for bonds if one studies the early postwar period of slowly rising inflation or the very recent period of spectacular stock returns.

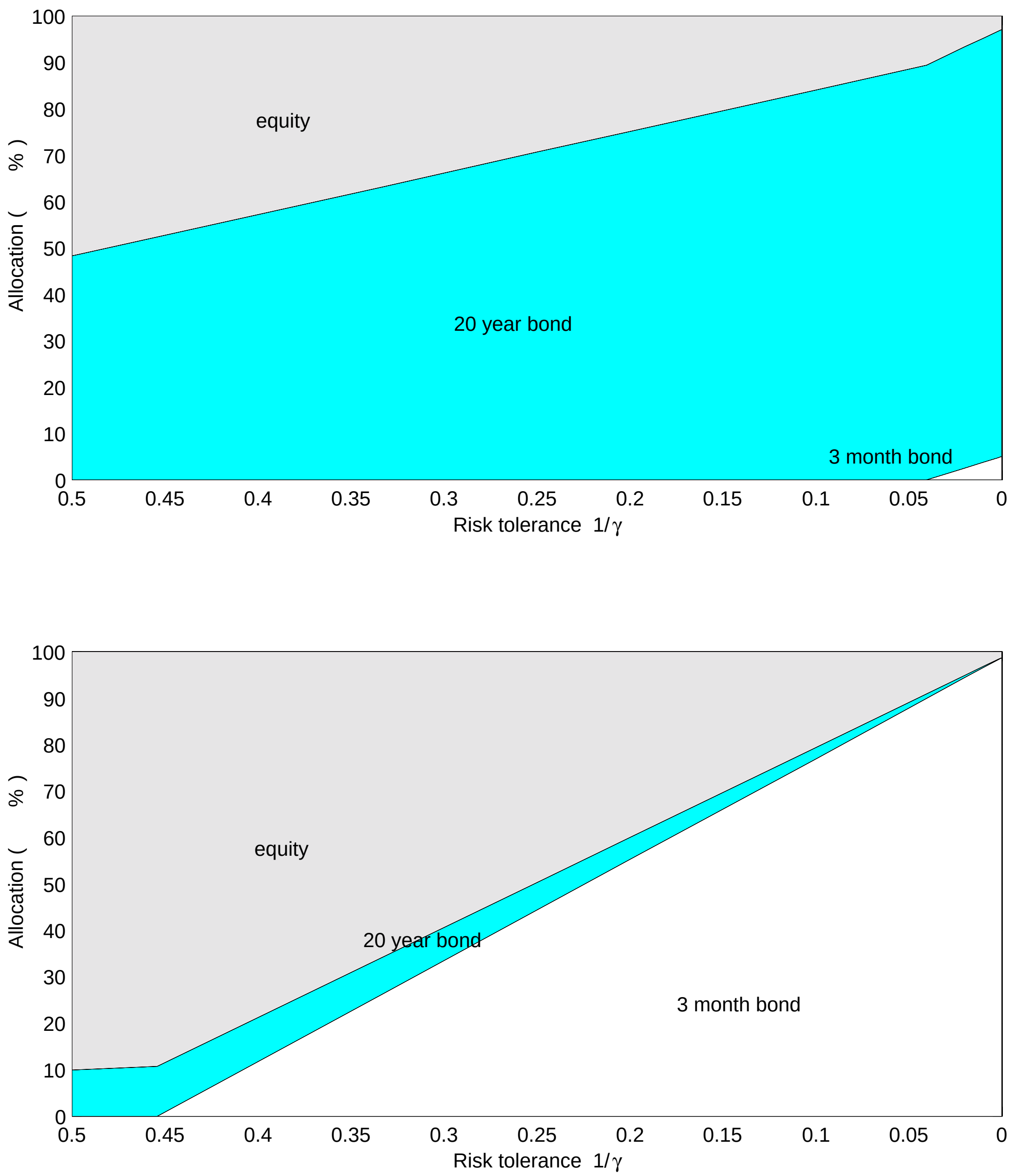

Figure 3.3: Optimal portfolios for US 1983-96

A long-horizon analysis treats bonds very differently, and assigns them a much more important role in the optimal portfolio. For long-term investors, money market investments are not riskless because they must be rolled over at uncertain future interest rates. Just as borrowers have come to appreciate that short-term debt carries a risk of having to refinance at high rates during a financial crisis, so long-term investors must appreciate that short-term investments carry the risk of having to reinvest at low real rates in the future. For long-term investors, an inflation-indexed long-term bond is actually less risky than cash. Such a bond does not have a stable market value in the short term, but it delivers a predictable stream of real income and thus supports a stable standard of living in the long term.

The implications for portfolio choice depend both on the assets that are available, and on the investor's view about the risk of inflation. If inflation-indexed bonds are available, then long-term investors should shift out of equities and into inflation-indexed bonds as they become more conservative. Even if only nominal bonds are available, conservative long-term investors should hold large positions in long-term bonds if they believe that inflation risk is low as we have estimated it to be in the US in the period 1983–96. In this sense the message of this chapter might be summarized as "Bonds, James, Bonds"!

Inflation risk is however a serious caveat. In the presence of significant inflation risk, of the sort we have estimated for the US in the period 1952–96, nominal bonds are risky assets for long-term investors and are not good substitutes for inflation-indexed bonds. This conclusion illustrates the general point that strategic asset demands depend on many features of the environment: not just on the conditional means and variances of returns that determine myopic asset demands, but also on the processes driving relevant state variables such as inflation and real interest rates.

Chapter 4

Is the Stock Market Safer for Long-Term Investors?

In Chapter 3 we studied strategic asset allocation in a model with time-varying real interest rates and inflation, and constant risk premia on all assets. We found that short-term safe assets are not riskless for long-term investors, because they must be rolled over at uncertain future rates. The riskless asset for a long-term investor is a long-term inflation-indexed bond, and nominal bonds are also close to riskless if inflation risk is low. Thus conservative long-term investors should tilt their portfolios towards bonds, rather than towards cash as predicted by the standard short-term analysis.

The model of Chapter 3 does not imply any special role for stocks in the portfolios of long-term investors. Intertemporal hedging demand in that model is determined by covariance with future real interest rates, which is the same for stocks as for bonds of the same duration; by variance, which is higher for stocks than for bonds; and by covariances among the available assets. Thus intertemporal hedging considerations benefit bonds, rather than stocks. An aggressive long-term investor will hold stocks because of their high average returns, but this is the same consideration that influences an aggressive short-term investor.

These results contrast with the commonly held view that long-term investors can afford to increase their stockholdings because stocks are comparatively safe for such investors. Jeremy Siegel (1994) expresses this view particularly clearly in his popular book *Stocks for the Long Run*:

> "It is widely known that stock returns, on average, exceed bonds in the long run. But it is little known that in the long run, the risks in stocks are *less than* those found in bonds or

> even bills!.... Real stock returns are substantially more volatile than the returns of bonds and bills over short-term periods. But as the horizon increases, the range of stock returns narrows far more quickly than for fixed-income assets.... Stocks, in contrast to bonds or bills, have never offered investors a negative real holding period return yield over 20 years or more. Although it might appear riskier to hold stocks than bonds, precisely the opposite is true: the safest long-term investment has clearly been stocks, not bonds." (pp. 29–30).

We saw in Chapter 2 that if asset returns are independent and identically distributed (iid) over time, then there is a precise mathematical relationship between risk at a short horizon and at a long horizon. Siegel's measure of risk is the standard deviation of the annualized return, which must be inversely proportional to the square root of the horizon if returns are iid. Any evidence that risk does not scale with horizon in this way is indirect evidence for predictability of asset returns. In order to evaluate such evidence we need a general empirical framework that allows for predictability—not just predictability of real interest rates, as in Chapter 3, but predictability of risk premia as well.

In this chapter we use a vector autoregressive (VAR) system to capture the historical predictability of asset returns. This type of model has been used in a similar context by Kandel and Stambaugh (1987), Campbell (1991, 1996), Hodrick (1992), and Barberis (2000) among others. Section 4.1 develops the framework and solution method, and section 4.2 applies it to measure stock and bond market risk in historical US data. Section 4.1.1 sets up the model, and section 4.1.2 extends our approximate solution method for the intertemporal consumption and portfolio choice problem to handle this case. Section 4.1.3 studies a special but illuminating case, solved in Campbell and Viceira (1999), in which there is a constant real interest rate and the investor allocates wealth between short-term safe assets and stocks, which follow a mean-reverting process. Section 4.2.1 estimates a VAR on historical US data, section 4.2.2 derives implications for asset risks at different investment horizons, and section 4.2.3 derives optimal portfolio weights for bills, nominal bonds, and stocks. In section 4.2.4 we use the VAR to impute hypothetical returns on inflation-indexed bonds, redoing the portfolio analysis for the case where inflation-indexed bonds are available. Section 4.3 concludes. The empirical work in the chapter is closely based on Campbell, Chan, and Viceira (2000).

4.1 Long-Term Portfolio Choice in a VAR Model

4.1.1 VAR specification

We begin by establishing notation that can handle multiple risky assets and forecasting variables. We work with a risky benchmark return, $r_{0,t+1}$, and a vector of n excess returns over the benchmark return, $\mathbf{r}_{t+1} - r_{0,t+1}\boldsymbol{\iota}$. We include other forecasting variables, such as the nominal interest rate or the dividend-price ratio on stocks, in a vector $\mathbf{s}_{t+1}$. We stack the benchmark return, excess risky returns, and other state variables into a single $m \times 1$ state vector $\mathbf{z}_{t+1}$:

$$\mathbf{z}_{t+1} \equiv \begin{bmatrix} r_{0,,t+1} \\ \mathbf{r}_{t+1} - r_{0,t+1}\boldsymbol{\iota} \\ \mathbf{s}_{t+1} \end{bmatrix}. \tag{4.1}$$

We postulate that the dynamics of $\mathbf{z}_{t+1}$ are well captured by a first-order vector autoregressive process or VAR(1). The use of a VAR(1) is not restrictive since any vector autoregression can be rewritten in this form through an expansion of the vector of state variables. Then we have

$$\mathbf{z}_{t+1} = \boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_1 \mathbf{z}_t + \mathbf{v}_{t+1}, \tag{4.2}$$

where $\boldsymbol{\Phi}_0$ is the $m \times 1$ vector of intercepts, $\boldsymbol{\Phi}_1$ is the $m \times m$ matrix of slope coefficients, and $\mathbf{v}_{t+1}$ is the $m \times 1$ vector of shocks to the state variables. We assume that $\mathbf{v}_{t+1}$ is normally distributed white noise, with mean zero and variance-covariance matrix Σ_v. Thus, we allow the shocks to be cross-sectionally correlated, but assume that they are homoskedastic and independently distributed over time. The VAR framework conveniently captures the dependence of expected returns of various assets on their past histories as well as on other predictive variables. The stochastic evolution of these other state variables $\mathbf{s}_{t+1}$ is also determined by the system.

The assumption of homoskedasticity is of course restrictive. It rules out the possibility that the state variables predict changes in risk; they can affect portfolio choice only by predicting changes in expected returns. Authors such as Campbell (1987), Harvey (1989, 1991), and Glosten, Jagannathan, and Runkle (1993) have explored the ability of the state variables used here to predict risk and have found only modest effects that seem to be dominated by the effects of the state variables on expected returns. In the next chapter, following Chacko and Viceira (1999), we show how to include changing risk in a long-term portfolio choice problem.

It is common in the continuous-time finance literature to assume that markets are complete, that is, that the state variables governing investment opportunities are driven by the same stochastic processes that drive asset returns so that innovations to investment opportunities are perfectly hedgeable using financial assets. The model (4.1) does not make this assumption. Whenever there are additional state variables $\mathbf{s}_{t+1}$ in the vector $\mathbf{z}_{t+1}$, and whenever the variance-covariance matrix of the VAR system Σ_v has full rank, then shocks to investment opportunities are imperfectly correlated with shocks to asset returns and cannot be perfectly hedged using financial assets. The ability to handle incomplete markets is an important empirical advantage of this model.

Given our homoskedastic VAR formulation, the unconditional distribution of $\mathbf{z}_t$ is easily derived. The state vector $\mathbf{z}_t$ inherits the normality of the shocks $\mathbf{v}_{t+1}$. It has unconditional mean $\boldsymbol{\mu}_z$ and variance-covariance matrix $\boldsymbol{\Sigma}_z$ that can straightforwardly be calculated from the VAR coefficients $\boldsymbol{\Phi}_0$, $\boldsymbol{\Phi}_1$, and $\boldsymbol{\Sigma}_v$. We can also calculate the conditional moments of linear and quadratic combinations of the variables.

4.1.2 Solving the model

We seek a solution that satisfies the loglinear Euler equations for the Epstein-Zin model, given the approximations laid out in the previous two chapters. That is, we need to find consumption and portfolio rules that satisfy the consumption Euler equation (2.46) and the portfolio Euler equation for multiple risky assets (2.48). The consumption Euler equation can be rewritten, using our loglinear approximation to the intertemporal budget constraint, as a difference equation in $c_t - w_t$:

$$c_t - w_t = -\rho\psi \log\delta - \rho\chi_{p,t} + \rho(1-\psi)\mathrm{E}_t(r_{p,t+1}) + \rho k + \rho\mathrm{E}_t(c_{t+1} - w_{t+1}), \quad (4.3)$$

where $\chi_{p,t} = (\theta/2\psi)\mathrm{Var}_t\left(\Delta c_{t+1} - \psi r_{p,t+1}\right)$. The portfolio Euler equation can be rewritten, using our approximation to the intertemporal budget constraint and the fact that $\Delta c_{t+1} = \Delta(c_{t+1} - w_{t+1}) + \Delta w_{t+1}$, as

$$E_t(\mathbf{r}_{t+1} - r_{0,t+1}\boldsymbol{\iota}) + \frac{\boldsymbol{\sigma}_t^2}{2} = \frac{\theta}{\psi}\boldsymbol{\sigma}_{c-w,t} + \gamma\boldsymbol{\sigma}_{p,t} - \boldsymbol{\sigma}_{0,t}, \quad (4.4)$$

where $\boldsymbol{\sigma}_t^2$ is a vector containing the variances of excess returns, $\boldsymbol{\sigma}_{c-w,t} = \mathrm{Cov}_t(\mathbf{r}_{t+1} - r_{0,t+1}\boldsymbol{\iota}, c_{t+1} - w_{t+1})$ is a vector containing the covariances of excess returns with the log consumption-wealth ratio, $\boldsymbol{\sigma}_{p,t} = \mathrm{Cov}_t(\mathbf{r}_{t+1} - r_{0,t+1}\boldsymbol{\iota}, r_{p,t+1})$ is a vector containing the covariances of excess returns with the log portfolio return, and $\boldsymbol{\sigma}_{0,t} = \mathrm{Cov}_t(\mathbf{r}_{t+1} - r_{0,t+1}\boldsymbol{\iota}, r_{0,t+1})$ is a vector containing the covariances of excess returns with the log return on the benchmark asset.

To solve the model, we now guess that the optimal portfolio and consumption rules take the form

$$\boldsymbol{\alpha}_t = \mathbf{a}_0 + \mathbf{A}_1 \mathbf{z}_t, \tag{4.5}$$

$$c_t - w_t = b_0 + \mathbf{b}_1' \mathbf{z}_t + \mathbf{b}_2' \mathbf{z}_t \mathbf{z}_t' \mathbf{b}_2. \tag{4.6}$$

That is, the optimal portfolio rule is linear in the VAR state vector but the optimal consumption rule is quadratic. $\mathbf{a}_0$ is an n-vector, $\mathbf{A}_1$ is an $n \times m$ matrix, b_0 is a scalar, and $\mathbf{b}_1$ and $\mathbf{b}_2$ are m-vectors.

The motivation for this guess is that (4.5) is the simplest portfolio rule that allows the investor to shift his portfolio in response to changing risk premia. All the variables in the state vector $\mathbf{z}_t$ can potentially affect risk premia, and thus the portfolio vector $\boldsymbol{\alpha}_t$ must be free to respond to them; we assume that it does so linearly. Given a linear portfolio rule, the expected portfolio return implied by (2.22) is quadratic in the state variables. State variables affect the expected portfolio return both directly, by shifting the expected returns on existing asset holdings, and indirectly, by shifting the asset allocations. Since each of these effects is linear, their interaction is quadratic and this makes the expected portfolio return quadratic. But then the consumption Euler equation (4.3) implies that the log consumption-wealth ratio must also be at least quadratic in the state variables. The consumption Euler equation also has a conditional variance term, but this too turns out to be quadratic in the state variables given our homoskedastic VAR specification. Kim and Omberg (1996) derived a similar linear-quadratic solution for a finite-horizon continuous-time model in which the investor has power utility defined over terminal wealth.

To verify our guess and solve for the parameters of the solution, we write the conditional moments that appear in (4.4) as functions of the VAR coefficients and the unknown parameters of (4.5) and (4.6). We then solve for the parameters that satisfy (4.4). This gives us the parameters in (4.5) as functions of the VAR coefficients and the still unknown parameters of (4.6). Next we substitute into (4.3), both sides of which are quadratic in the VAR state variables given our conjectured quadratic form for the optimal consumption-wealth ratio. Finally we solve this system of quadratic equations for the parameters of (4.6).

Given the loglinearization parameter ρ, this solution is analytical. Campbell and Viceira (1999) write it out explicitly for the special case in which there is a constant riskless interest rate, a single risky asset, and a single forecasting variable for the excess risky return which itself follows an AR(1) process. We explain this case in section 4.1.3. Campbell, Chan, and Viceira (2000) extend the approach to the general VAR case, for which it is more convenient to solve the linear-quadratic equations numerically.

The solution presented here is exact in continuous time when asset prices follow continuous diffusion processes, if the consumption-wealth ratio is constant. The consumption-wealth ratio is constant if the elasticity of intertemporal substitution $\psi = 1$, but in all other cases, the solution is only an approximation. Campbell and Koo (1997) and Campbell, Cocco, Gomes, Maenhout, and Viceira (1999) evaluate the accuracy of the approximate solution, respectively in models with exogenous portfolio returns and exogenous returns on underlying assets, and find that it is acceptably accurate for elasticities of intertemporal substitution up to about 3. In particular, this implies that low elasticities of intertemporal substitution of the sort estimated by macroeconomists should be consistent with the use of the approximate solution.

The case $\psi = 1$ is also important because only in this case do we know that the value of ρ equals δ. In all other cases we must solve for ρ along with the other parameters of the model. We do this numerically; we initialize $\rho = \delta$, solve for the other parameters of the model, calculate the implied mean log consumption-wealth ratio from (4.6), recalculate ρ, and repeat until convergence. This process is extremely rapid except in cases where the infinite-horizon optimization problem is ill-defined (for example, because average rates of return are too high relative to the investor's rate of time preference, so that a finite-horizon investor's utility diverges as the investment horizon increases).

An important property of the model is that given the loglinearization parameter ρ, the optimal portfolio rule does not depend on the intertemporal elasticity of substitution ψ. ψ only affects portfolio choice to the extent that it enters into the determination of ρ. Empirically, this indirect effect through ρ seems to be minor.

4.1.3 A special case: One risky asset and a constant real interest rate

Campbell and Viceira (1999) study a special case in which there is a short-term riskless asset with a constant real log return r_f. Because this riskless real return is constant, it is a safe investment both for short-term investors and for long-term investors. Thus the issue emphasized in Chapter 3, that the identity of the riskless asset may be different for investors with different horizons, does not arise here. Campbell and Viceira (1999) also assume that there is a single risky asset ("stocks") with log return r_{t+1} given by

$$r_{t+1} - \mathrm{E}_t r_{t+1} = u_{t+1}, \tag{4.7}$$

where u_{t+1} is the innovation to the risky asset return, normally distributed with mean zero and variance σ_u^2. The expected excess log return on the risky asset, adjusted for one-half its variance in the familiar manner, equals a state variable x_t:[1]

$$\mathrm{E}_t r_{t+1} - r_f + \frac{\sigma_u^2}{2} = x_t. \tag{4.8}$$

Finally, x_t follows an AR(1) process with mean μ and persistence ϕ. The innovation to x_{t+1} is written η_{t+1}, assumed to be normally distributed with mean zero and variance σ_η^2:

$$x_{t+1} = \mu + \phi(x_t - \mu) + \eta_{t+1}. \tag{4.9}$$

Modelling mean reversion

The innovations u_{t+1} and η_{t+1} can be correlated, with covariance $\sigma_{\eta u}$. In fact, this covariance is what generates intertemporal hedging demand for the risky asset by long-term investors. The state variable x_t summarizes investment opportunities at time t. Thus the conditional covariance between the risky asset return and the state variable measures the ability of the risky asset to hedge time-variation in investment opportunities. This covariance is given by

$$\mathrm{Cov}_t(r_{t+1}, x_{t+1}) = \mathrm{Cov}_t(r_{t+1}, r_{t+2}) = \sigma_{\eta u}. \tag{4.10}$$

The model can be solved for arbitrary $\sigma_{\eta u}$, but we focus attention on the case where $\sigma_{\eta u} < 0$. This case captures the notion that stocks are "mean-reverting"; an unexpectedly high return today reduces expected returns in the future, and thus high short-term returns tend to be offset by lower returns over the long term. This offset reduces the conditional variance of long-term stock returns, since

$$\begin{aligned} \mathrm{Var}_t(r_{t+1} + r_{t+2}) &= 2\mathrm{Var}_t(r_{t+1}) + 2\mathrm{Cov}_t(r_{t+1}, r_{t+2}) \\ &= 2\mathrm{Var}_t(r_{t+1}) + 2\sigma_{\eta u} < 2\mathrm{Var}_t(r_{t+1}). \end{aligned} \tag{4.11}$$

[1] This is a slight change from the parameterization of Campbell and Viceira (1999), which omitted the term $\sigma_u^2/2$ on the left-hand side of (4.8). Accordingly some equations here are slightly altered from the corresponding equations in Campbell and Viceira.

That is, the conditional variance of stock returns does not grow in proportion with the investment horizon, but grows more slowly. If we calculate a conditional variance ratio,

$$VR_t(K) \equiv \frac{\mathrm{Var}_t(r_{t+1} + r_{t+2} + ...r_{t+K})}{K\mathrm{Var}_t(r_{t+1})}, \tag{4.12}$$

the ratio will be less than one at all horizons K. Stocks will appear relatively safer to long-term investors.

The discussion above concentrates on conditional variances, since these are what matter to investors. The empirical literature on mean-reversion, initiated by Poterba and Summers (1988) and Fama and French (1988b), typically emphasizes unconditional variances instead. These can behave quite differently from conditional variances since

$$\mathrm{Cov}(r_{t+1}, x_{t+1}) = \mathrm{Cov}(x_t + u_{t+1}, x_{t+1}) = \phi\sigma_x^2 + \sigma_{\eta u}. \tag{4.13}$$

Persistence in the process for x_{t+1} can make this unconditional covariance zero, even if the conditional covariance is negative. Campbell (1991) and Campbell, Lo, and MacKinlay (1997, Chapter 7) emphasize this point. More generally, the unconditional variance ratio,

$$VR(K) \equiv \frac{\mathrm{Var}(r_{t+1} + r_{t+2} + ...r_{t+K})}{K\mathrm{Var}(r_{t+1})} = \frac{VR_t(K)}{1 - R^2(K)}, \tag{4.14}$$

where $R^2(K)$ is the explanatory power of a regression of the K-period stock return onto the state variable x_t. Thus the unconditional variance ratio is always greater than the conditional variance ratio; empirical results using the former understate the risk-reduction that is relevant for long-term investors. The difference between the two variance ratios can be substantial, since authors such as Fama and French (1988a) and Campbell and Shiller (1988a,b) have found long-horizon R^2 statistics as large as 40%. We explore this issue empirically in section 4.2.2.

Solving the model

The model is a special case of the general VAR system. Thus the solution takes the form

$$\alpha_t = a_0 + a_1 x_t, \tag{4.15}$$

$$c_t - w_t = b_0 + b_1 x_t + b_2 x_t^2. \tag{4.16}$$

The quadratic expression for the consumption-wealth ratio implies that the value function in (2.49) takes the exponential-quadratic form

$$V_t = \exp\left[\frac{b_0 - \psi \log(1-\delta)}{1-\psi} + \frac{b_1}{1-\psi}x_t + \frac{b_2}{1-\psi}x_t^2\right]. \tag{4.17}$$

Kim and Omberg (1996) derive a similar exponential-quadratic solution in a related continuous-time model, where a single state variable follows a continuous-time AR(1) (Ornstein-Uhlenbeck) process and the investor has power utility defined over terminal wealth.

Given the form of the value function, it should not be surprising that the ratios $b_1/(1-\psi)$ and $b_2/(1-\psi)$ play a key role in the solution. These ratios capture the linear and quadratic effects of the state variable x_t on utility, which have the same sign as the effects of x_t on consumption only when income effects dominate substitution effects, that is, when $\psi < 1$. Campbell and Viceira (1999) show that the parameters of the portfolio rule are related to $b_1/(1-\psi)$ and $b_2/(1-\psi)$ as follows:

$$a_0 = \left(1 - \frac{1}{\gamma}\right)\left[\left(\frac{b_1}{1-\psi}\right) + 2\mu(1-\phi)\left(\frac{b_2}{1-\psi}\right)\right]\left(-\frac{\sigma_{\eta u}}{\sigma_u^2}\right), \tag{4.18}$$

$$a_1 = \frac{1}{\gamma\sigma_u^2} + \left(1 - \frac{1}{\gamma}\right)\left[2\phi\left(\frac{b_2}{1-\psi}\right)\right]\left(-\frac{\sigma_{\eta u}}{\sigma_u^2}\right). \tag{4.19}$$

For a myopic investor, we would have $a_0 = 0$ and $a_1 = 1/\gamma\sigma_u^2$, so intertemporal hedging demand accounts for the entire right-hand side of (4.18) and the second term on the right-hand side of (4.19).

Campbell and Viceira (1999) show that $b_2/(1-\psi) > 0$ and does not depend on the average excess stock return μ, while $b_1/(1-\psi) = 0$ when $\mu = 0$ and should be expected to have the same sign as μ. The empirically relevant case is that where μ, $b_1/(1-\psi)$, and $b_2/(1-\psi)$ are all positive, while $\sigma_{\eta u}$ is negative so $-\sigma_{\eta u}/\sigma_u^2$ is also positive. In this case the intercept of the portfolio rule, a_0, is positive for conservative investors with $\gamma > 1$, which implies that such investors will hold stocks even when the expected excess return is zero. This is a striking result since it contradicts the well-known principle of short-term portfolio choice that a risk-averse investor will never wish to take on risk without receiving a reward for it.

The explanation for this result is as follows. With $\sigma_{\eta u} < 0$, stocks tend to have high returns when their expected future returns fall. With $\mu > 0$, the investor is normally long in stocks, so a decline in expected future stock returns is normally a deterioration in the investment opportunity set. A conservative investor with $\gamma > 1$ wants to hedge the risk of deteriorating investment opportunities by holding assets that deliver increased wealth when investment opportunities are poor. Stocks are just such an asset, so the investor has positive hedging demand for stock even when the current risk premium on stocks is zero.

Although the investor is normally long in stocks, if the expected excess return becomes significantly negative, a decline in expected future stock returns can represent an improvement in the investment opportunity set because it creates a profitable opportunity to short stocks. At this point in the state space the sign of hedging demand for stocks reverses. Hedging demand thus moves in the same direction as the state variable x_t, which explains why the slope of hedging demand—the second term on the right-hand side of (4.19)—is positive. The positive slope of hedging demand allows it to reverse sign for sufficiently negative x_t.[2] Perhaps surprisingly, the positive slope of hedging demand implies that long-term investors are more aggressive market timers than myopic investors, although the difference in slope is quite modest for empirically reasonable parameter values.

This solution is illustrated in Figure 4.1, which shows three alternative portfolio rules for a conservative investor with $\gamma > 1$ facing the benchmark case of mean-reverting stock returns. The horizontal line is a buy-and-hold strategy that assumes a constant expected excess stock return equal to the true unconditional mean μ. The line marked "Myopic investor" is the optimal strategy for a single-period investor who observes the conditional expected stock return x_t. The line marked "Strategic investor" is the optimal strategy for a long-term investor. This line is slightly steeper than the others, and is shifted upwards so that it has a positive intercept.

The discussion so far assumes that the average level of excess stock returns, μ, is positive. Positive average excess stock returns lead the investor normally to maintain a long position in stocks for which a decrease in the expected stock return represents a deterioration in investment opportunities. If μ were negative, however, the investor would normally have a short position in stocks for which a decrease in the expected stock return represents an improvement in investment opportunities. In this case the normal sign of hedging demand would be negative for an investor with $\gamma > 1$. The slope of hedging demand is unaffected by the coefficient μ, however, so in this case a sign reversal of the normal hedging demand occurs for sufficiently positive x_t. The case $\mu = 0$ is intermediate; in this case we have a symmetrical model in which the investor gains equally from increases and decreases in x_t away from its mean, and hedging demand has a positive slope but no

[2] Kim and Omberg [1996] give a clear account of this effect (Figure 4 and pp. 153–154).

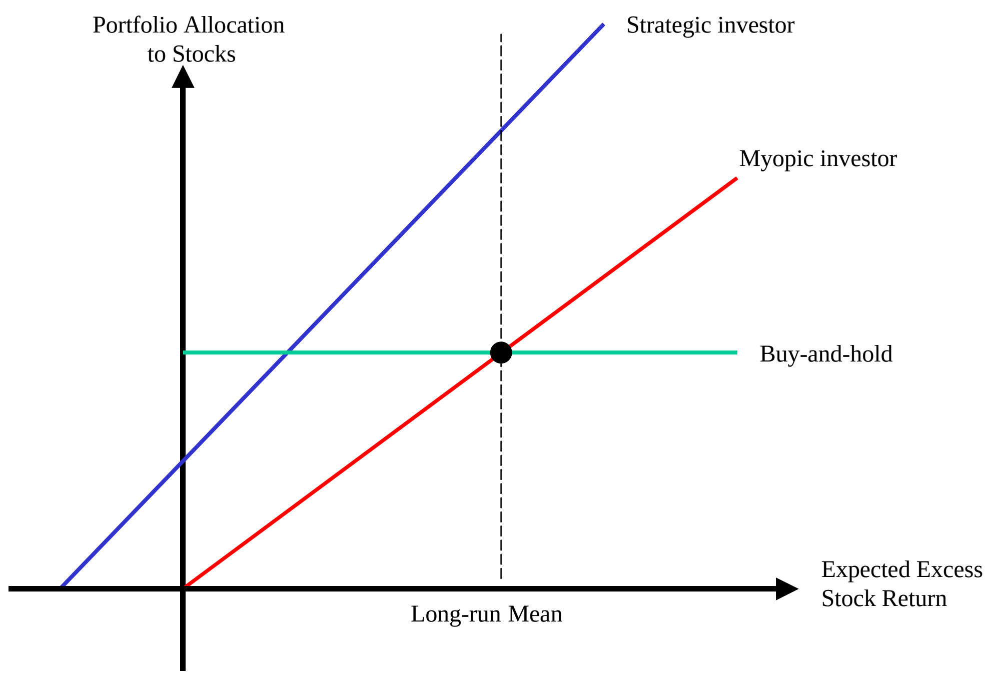

Figure 4.1: Alternative portfolio rules

intercept.

The model implies that hedging demand is not monotonic in risk aversion. At $\gamma = 1$, hedging demand is zero so initially it increases with risk aversion in the benchmark case. Eventually, however, hedging demand decreases with risk aversion since a sufficiently conservative investor will only hold the safe asset; such an investor will not wish to exploit stock-market opportunities and will not have any motive to hedge variation in those opportunities. One can show that as γ increases, the coefficients $b_1/(1-\psi)$, $b_2/(1-\psi)$, a_0, and a_1 all approach zero.

Campbell and Viceira (1999) examine the empirical implications of this model in some detail, using the log dividend-price ratio as an empirical proxy for the state variable x_t. Unfortunately there is an error in the empirical estimates reported in that paper; the estimates understate the predictability of postwar quarterly stock returns and the absolute value of the correlation between innovations in stock returns and dividend yields, and thus understate the magnitude of hedging demands. This error is explained in Campbell and Viceira (2000), which reports corrected results.

It is interesting to relate the model of this section to the advice of Siegel (1994) that long-term investors should aggressively buy and hold equities. Siegel bases his advice on the reduced risk of stock returns at long horizons. We have seen that such reduced risk can only arise from mean reversion in stock returns, a fact recognized by Siegel when he writes:

> "Stocks have what economists call mean-reverting returns, meaning that over long periods of time, high returns seem to be followed by periods of low returns and vice versa. On the other hand, over time, real returns on fixed-income assets become relatively less certain. For horizons of 20 years or more, bonds are riskier than stocks." (p.33).

The difficulty with Siegel's investment advice is that mean reversion is equivalent to predictable variation in stock returns, and such predictable variation is inconsistent with the optimality of a buy-and-hold strategy. The optimal strategy is instead a strategic market-timing strategy of the sort illustrated in Figure 4.1. Campbell and Viceira (1999, 2000) show that there are large utility losses from ignoring the market-timing aspect of the optimal investment strategy. Siegel's investment advice is clearly suboptimal unless an investor is constrained from taking on leverage, in which case the constrained optimal strategy might involve a 100% equity allocation over much of the state space. Even in this case, however, sufficiently positive past returns would drive down the expected future return to the point where the long-term investor should cut back his equity allocation.

4.2 Stock and Bond Market Risk in Historical US Data

4.2.1 Data and VAR estimation

In the previous section we provided a general theoretical framework for strategic asset allocation, and explored a special case with a constant real interest rate and a mean-reverting stock return. Although that case is illuminating, it is too special to provide a solid foundation for investment advice. In this section, we return to the general framework and use it to investigate how investors who differ in their consumption preferences and risk aversion allocate their portfolios among three assets: stocks, nominal bonds, and nominal Treasury bills. Investment opportunities are described by a VAR system that includes short-term ex-post real interest rates, excess stock returns, excess bond returns and variables that have been identified as return predictors by empirical research: the short-term nominal interest rate, the dividend-price ratio, and the yield spread between long-term bonds and Treasury bills.

The short-term nominal interest rate has been used to predict stock and bond returns by authors such as Fama and Schwert (1977), Campbell (1987), and Glosten, Jagannathan, and Runkle (1993). An alternative approach, suggested by Campbell (1991) and Hodrick (1992), is to stochastically detrend the short-term rate by subtracting a backwards moving average (usually measured over one year). For two reasons we do not adopt this alternative here. First, we emphasize a long-term annual data set in which we cannot measure a one-year moving average of short rates. Second, we want our model to capture inflation dynamics. If we include both the ex-post real interest rate and the nominal interest rate in the VAR system, we can easily calculate inflation by subtracting one from the other. This allows us to separate nominal from real variables, so that we can extend our model to include a hypothetical inflation-indexed bond in the menu of assets. We consider this extension in section 4.2.4.

We compute optimal portfolio rules for different values of γ, assuming either $\psi = 1$ or $\psi = \gamma^{-1}$. In both cases, we set $\delta = 0.92$ in annual terms. The first case gives the exact solution of Giovannini and Weil (1989), where the consumption-wealth ratio is constant and equal to $1 - \delta$. This implies that the loglinearization parameter $\rho \equiv 1 - \exp(\mathrm{E}[c_t - w_t])$ is equal to δ. The second case is the familiar power utility specification.

Data description

Our calibration exercise is based on annual and quarterly data for the US stock market. The annual dataset covers over a century from 1890 to 1995. Its source is the data used in Grossman and Shiller (1981), updated for the recent period by Campbell (1999).[3] This dataset contains data on prices and dividends on S&P 500 stocks as well as data on inflation and short-term interest rates. The equity price index is the end-of-December S&P 500 Index, and the price index is the Producer Price Index. The short rate is the return on 6-month commercial paper bought in January and rolled over July. We use this dataset to construct time series of short-term, nominal and ex-post real interest rates, excess returns on equities, and dividend-yields. Finally, we obtain data on long-term nominal bonds from the long yield series in Shiller (1989), which we have updated using the Moody's AAA corporate bond yield average. We construct the long bond return from this series using the loglinear approximation technique described in Chapter 10 of Campbell, Lo and MacKinlay (1997):

$$r_{n,t+1} \approx D_{n,t} y_{n,t} - (D_{n,t} - 1)\, y_{n-1,t+1},$$

where n is bond maturity, the bond yield is written Y_{nt}, the log bond yield $y_{n,t} = \log(1 + Y_{n,t})$, and $D_{n,t}$ is bond duration. We calculate duration at time t as

$$D_{n,t} \approx \frac{1 - (1 + Y_{n,t})^{-n}}{1 - (1 + Y_{n,t})^{-1}},$$

and we set n to 20 years. We also approximate $y_{n-1,t+1}$ by $y_{n,t+1}$.

The quarterly data begin in 1952:2, shortly after the Fed-Treasury Accord that fundamentally changed the stochastic process for nominal interest rates, and end in 1997:4. We obtain our quarterly data from the Center for Research in Security Prices (CRSP). We construct the ex post real Treasury bill rate as the difference of the log return (or yield) on a 90-day bill and log inflation, and the excess log stock return as the difference between the log return on a stock index and the log return on the 90-day bill. We use the value-weighted return, including dividends, on the NYSE, NASDAQ and AMEX markets. We construct the excess log bond return in a similar

[3] See the Data Appendix to Campbell (1999), available on the author's website.

way, using the 5-year bond return from the US Treasury and Inflation Series (CTI) file in CRSP.

The nominal yield on Treasury bills is the log yield on a 90-day bill. To calculate the dividend-price ratio, we first construct the dividend payout series using the value-weighted return including dividends, and the price index series associated with the value-weighted return excluding dividends. Following the standard convention in the literature, we take the dividend series to be the sum of dividend payments over the past year. The dividend-price ratio is then the log dividend less the log price index. The yield spread is the difference between the 5-year zero-coupon bond yield from the CRSP Fama-Bliss data file (the longest yield available in the file) and the bill rate.

VAR estimation

Table 4.1 gives the first and second sample moments of the data. Except for the dividend-price ratio, the sample statistics are in annualized, percentage units. Mean excess log returns are adjusted by one-half their variance to account for Jensen's Inequality. For the postwar quarterly dataset, Treasury bills offer a low average real return (a mere 1.813% per year) along with low variability. Stocks have an excess return of 7.119% per year compared to 0.712% for the 5-year bond. Although volatility is much higher for stocks than for bonds (16.093% vs. 5.576%), the Sharpe ratio is almost three and a half times as high for stocks as for bonds. The average Treasury bill rate and yield spread are 5.867% and 0.616%, respectively.

Covering a century of data, the annual dataset gives a different description of the relative performance of each asset. The real return on short-term nominal debt is quite volatile, due to greater volatility in both real interest rates and inflation before World War II. Stocks offer a slightly lower excess return, and yet a higher standard deviation, than the postwar quarterly data. The Depression period is largely responsible for this result. The long-term bond also performs rather poorly, giving a Sharpe ratio of only 0.105 versus a Sharpe ratio of 0.345 for stocks. The bill rate has a lower mean in the annual dataset, but the yield spread has a higher mean. Both bill rates and yield spreads have higher standard deviations in the annual dataset. Figure 2 plots the history of the variables included in the annual VAR.

Table 4.2 reports the estimation results for the VAR system in the annual dataset, while Table 4.3 reports results for the quarterly dataset (Panel B). The top section of each table reports coefficient estimates (with t-statistics in parentheses) and the R^2 statistic (with the p-value of the F test of joint

Table 4.1: Sample Statistics

	Sample Moment	1890 - 1995	1952Q2 - 1997Q4
(1)	$E[r^{\$}_{1,t} - \pi_t] + \sigma^2(r^{\$}_{1,t} - \pi_t)/2$	2.112	1.813
(2)	$\sigma(r^{\$}_{1,t} - \pi_t)$	8.891	1.460
(3)	$E[r^{\$}_{e,t} - r^{\$}_{1,t}] + \sigma^2(r^{\$}_{e,t} - r^{\$}_{1,t})/2$	6.242	7.119
(4)	$\sigma(r^{\$}_{e,t} - r^{\$}_{1,t})$	18.107	16.093
(5)	$SR = (3)/(4)$	0.345	0.442
(6)	$E[r^{\$}_{n,t} - r^{\$}_{1,t}] + \sigma^2(r^{\$}_{n,t} - r^{\$}_{1,t})/2$	0.661	0.712
(7)	$\sigma(r^{\$}_{n,t} - r^{\$}_{1,t})$	6.299	5.576
(8)	$SR = (6)/(7)$	0.105	0.128
(9)	$E[y^{\$}_t]$	4.321	5.867
(10)	$\sigma(y^{\$}_t)$	2.611	1.555
(11)	$E[d_t - p_t]$	-3.079	-3.371
(12)	$\sigma(d_t - p_t)$	0.275	0.244
(13)	$E[y^{\$}_{n,t} - y^{\$}_{1,t}]$	0.876	0.616
(14)	$\sigma(y^{\$}_{n,t} - y^{\$}_{1,t})$	1.459	0.588

Note: $r^{\$}_{1,t}$ = log return on T-bills, π_t = log inflation rate, $r^{\$}_{e,t}$ = log return on equities, $r^{\$}_{n,t}$ = log return on nominal bond, $(d-p)_t$ = log dividend-price ratio, rb_t = relative bill rate, $y^{\$}_{n,t}$ = log yield on the nominal bond, and $y^{\$}_{1,t}$ is the short yield. The bond is a 5-year nominal bond in the quarterly dataset and a 20-year for the annual dataset.

significance in parentheses) for each equation in the system.[4] The bottom section of each table shows the covariance structure of the innovations in VAR system. The entries above the main diagonal are correlation statistics, and the entries on the main diagonal are standard deviations multiplied by 100. All variables in the VAR are measured in natural units, so standard deviations are per year in Table 4.2 and per quarter in Table 4.3.

The first row of each table corresponds to the real bill rate equation. Only the lagged real bill rate and the lagged nominal bill rate have t-statistics above 2 in both sample periods. The rest of the variables are either not significant or only marginally significant in predicting real bill rates one period ahead.

The second row corresponds to the equation for the excess stock return. Predicting excess stock returns is difficult: This equation has the lowest R^2 in the annual sample, and the second lowest R^2 in the quarterly sample. The dividend-price ratio, with a positive coefficient, is the only variable with a t-statistic well above 2. The coefficient on the lagged nominal short-term interest rate is marginally significant in the quarterly sample, and it has a negative sign in both samples. Lagged excess bond returns and yield spreads both have positive coefficients, but they are not statistically significant.

The third row is the equation for the excess bond return. In the long annual dataset, lagged excess returns on stocks and bonds, real Treasury bill rates, and yield spreads help predict future excess bond returns. In the quarterly postwar data, only lagged excess returns on stocks help predict future excess bond returns. The fit of the equation is also much worse than the fit in the annual sample. In part, this difference in results may reflect approximation error in our procedure for constructing annual bond returns; the possibility of such error should be kept in mind when interpreting our annual results.

The last three rows report the estimation results for the remaining state variables, each of which are fairly well described by a univariate AR(1) process. The nominal bill rate in the fourth row is predicted by the lagged yield spread in the quarterly data set, but the main predictor is the lagged nominal yield, whose coefficient is above 0.9 in both samples, implying extremely persistent dynamics. The log dividend-price ratio in the fifth row also has persistent dynamics; the lagged dividend-price ratio has a coefficient of 0.78

[4]We estimate the VAR imposing the restriction that the unconditional means of the variables implied by the VAR coefficient estimates equal their full-sample arithmetic counterparts. Standard, unconstrained least-squares fits exactly the mean of the variables in the VAR excluding the first observation. We use constrained least-squares to ensure that we fit the full-sample means.

Table 4.2: VAR Estimation Results
Annual Sample (1890 - 1995)

Dependent Variable	rtb_t (t)	xr_t (t)	xb_t (t)	y_t (t)	$(d-p)_t$ (t)	spr_t (t)	R^2 (p)
Coefficient Estimates							
rtb_{t+1}	0.305	-0.056	0.147	0.685	-0.004	-0.869	0.239
	(2.456)	(-1.391)	(0.973)	(2.271)	(-0.138)	(-1.320)	(0.000)
xr_{t+1}	0.111	0.087	-0.219	-0.134	0.187	1.199	0.086
	(0.420)	(0.703)	(-0.763)	(-0.191)	(3.449)	(0.908)	(0.117)
xb_{t+1}	0.200	0.095	-0.091	-0.083	0.009	2.573	0.421
	(3.153)	(2.658)	(-0.652)	(-0.238)	(0.474)	(5.070)	(0.000)
y_{t+1}	-0.042	-0.011	0.029	0.918	-0.005	-0.029	0.783
	(-1.928)	(-1.571)	(0.907)	(12.378)	(-1.027)	(-0.228)	(0.000)
$(d-p)_{t+1}$	-0.562	-0.134	0.529	-0.509	0.779	-1.686	0.677
	(-2.243)	(-1.228)	(1.937)	(-0.795)	(13.666)	(-1.255)	(0.000)
spr_{t+1}	0.019	0.002	-0.016	0.085	0.004	0.838	0.542
	(1.109)	(0.357)	(-0.709)	(1.633)	(1.093)	(8.655)	(0.000)

Cross-Correlation of Residuals	rtb	xr	xb	y	$(d-p)$	spr
rtb	7.753	-0.194	-0.029	0.130	0.131	-0.167
xr	-	17.303	0.027	-0.175	-0.713	0.210
xb	-	-	4.794	-0.636	-0.115	0.266
y	-	-	-	1.217	0.221	-0.903
$(d-p)$	-	-	-	-	15.624	-0.192
spr	-	-	-	-	-	0.987

Table 4.3: VAR Estimation Results
Quarterly Sample (1952Q2 - 1997Q4)

Dependent Variable	rtb_t (t)	xr_t (t)	xb_t (t)	y_t (t)	$(d-p)_t$ (t)	spr_t (t)	R^2 (p)
Coefficient Estimates							
rtb_{t+1}	0.445	0.006	-0.018	0.287	-0.001	0.034	0.324
	(6.142)	(1.040)	(-0.822)	(3.695)	(-0.519)	(0.207)	(0.000)
xr_{t+1}	0.256	0.061	0.355	-1.852	0.077	2.072	0.109
	(0.275)	(0.769)	(1.492)	(-2.097)	(3.011)	(0.857)	(0.001)
xb_{t+1}	0.159	-0.054	-0.050	0.310	0.003	0.995	0.043
	(0.447)	(-2.594)	(-0.424)	(0.771)	(0.424)	(0.949)	(0.179)
y_{t+1}	0.001	0.003	-0.008	0.962	0.000	0.490	0.796
	(0.016)	(0.990)	(-0.418)	(16.990)	(0.361)	(4.432)	(0.000)
$(d-p)_{t+1}$	-0.626	-0.064	-0.334	1.047	0.939	-1.926	0.892
	(-0.631)	(-0.744)	(-1.280)	(1.144)	(34.309)	(-0.783)	(0.000)
spr_{t+1}	-0.015	0.000	0.012	0.019	-0.001	0.497	0.277
	(-0.369)	(0.072)	(0.961)	(0.491)	(-0.876)	(6.460)	(0.000)

Cross-Correlation of Residuals

	rtb	xr	xb	y	$(d-p)$	spr
rtb	0.599	0.229	0.451	-0.511	-0.241	0.404
xr	-	7.586	0.274	-0.211	-0.969	0.108
xb	-	-	2.725	-0.766	-0.321	0.413
y	-	-	-	0.350	0.253	-0.899
$(d-p)$	-	-	-	-	7.961	-0.137
spr	-	-	-	-	-	0.250

in the annual data and 0.94 in the quarterly data. The yield spread in the sixth row also seems to follow an AR(1) process, but is considerably less persistent than the other variables, especially in the quarterly sample.

The bottom section of each table describes the covariance structure of the innovations in the VAR system. Unexpected log excess stock returns are highly negatively correlated with shocks to the log dividend-price ratio in both samples. This result is consistent with previous empirical results in Campbell (1991), Campbell and Viceira (1999), Stambaugh (1999) and others. Unexpected log excess bond returns are negatively correlated with shocks to the nominal bill rate, but positively correlated with the yield spread. This positive correlation is 41% in the quarterly sample, and about 27% in the annual sample.

The signs of these correlations help to explain the contrasting results of recent studies that apply Monte Carlo analysis to judge the statistical evidence for predictability in stock and bond returns. Stock-market studies typically find that asymptotic tests overstate the evidence for predictability of stock returns (Hodrick 1992, Goetzmann and Jorion 1993, Nelson and Kim 1993). Bond-market studies, on the other hand, find that asymptotic procedures are actually conservative and understate the evidence for predictability of bond returns (Bekaert, Hodrick, and Marshall 1997). The reason for the discrepancy is that asymptotic results in the stock market are based on positive regression coefficients of stock returns on the dividend-price ratio, while asymptotic results in the bond market are based on positive regression coefficients of bond returns on the yield spread. Stambaugh (1999) shows that the small-sample bias in such regressions has the opposite sign to the sign of the correlation between innovations in returns and innovations in the predictive variable. In the stock market the log dividend-price ratio is negatively correlated with returns, leading to a positive small-sample bias which helps to explain some apparent predictability; in the bond market, on the other hand, the yield spread is positively correlated with returns, leading to a negative small-sample bias which cannot explain the positive regression coefficient found in the data.

The signs of these correlations also have important effects on the volatility of bond and stock returns over long holding periods. We now explore these effects in some detail as they are highly relevant for long-term investors.

4.2.2 Return volatility at short and long horizons

Our estimated VAR system implies that there are important horizon effects on the relative volatilities of different investment strategies. In Figure 4.2

we plot the annualized standard deviations of real returns on stocks and bills implied by our annual VAR for investment horizons up to 100 years (panel A) and implied by our quarterly VAR for investment horizons up to 100 quarters or 25 years (panel B).[5] These standard deviations are conditional; that is, they take out movements that are predictable in advance and thus represent variation in investment opportunities rather than risk. We also plot conditional standard deviations for two alternative investment strategies using nominal bonds. The "long bond rolled" strategy keeps the maturity of the long bond constant at 20 years, buying a 20-year bond each year and selling it the next year in order to invest in a new 20-year bond. This is the strategy implicitly assumed in virtually all time series of long-term bond returns. The "bond held to maturity" strategy assumes that an investor with horizon k buys a nominal bond with k years to maturity and holds it until maturity. The standard deviation of the real return on this strategy is just the standard deviation of cumulative inflation from time t to time $t+k$, since a nominal bond held to maturity is riskless in nominal terms.

Figure 4.2 shows that stocks are mean-reverting—their long-horizon returns are less volatile than their short-horizon returns—while bonds and bills are mean-averting—their long-horizon returns are actually more volatile than their short-horizon returns. Mean-aversion is particularly strong for bills in the annual dataset, where persistent variation in the real interest rate amplifies the volatility of returns over long horizons. Mean-aversion also affects the returns on rolling long bonds in the annual dataset (because of both variation in the real interest rate and predictability of bond returns from the yield spread), and the returns on holding bonds to maturity in the quarterly dataset (because of persistent movements in inflation). Mean-reversion in stock returns was pointed out by Fama and French (1988b) and Poterba and Summers (1988), and has been the subject of much subsequent research. Siegel (1998) has used long-term data to directly measure mean-aversion in fixed-income securities and has emphasized its importance for long-term investors, but this phenomenon has received relatively little attention in the academic literature.

The estimated horizon effects on volatility are large enough to alter the rankings of asset return volatilities across investment horizons. In the annual system, stocks are far more volatile than bonds and bills at short horizons, but safer than bills or rolling bonds at long horizons, a point stressed by

[5] Note that we are not looking directly at the long-horizon properties of returns, but at the long-horizon properties of returns imputed from our first-order VAR. Thus, provided that our VAR captures adequately the dynamics of the data, we can consistently estimate the moments of returns over any desired horizon.

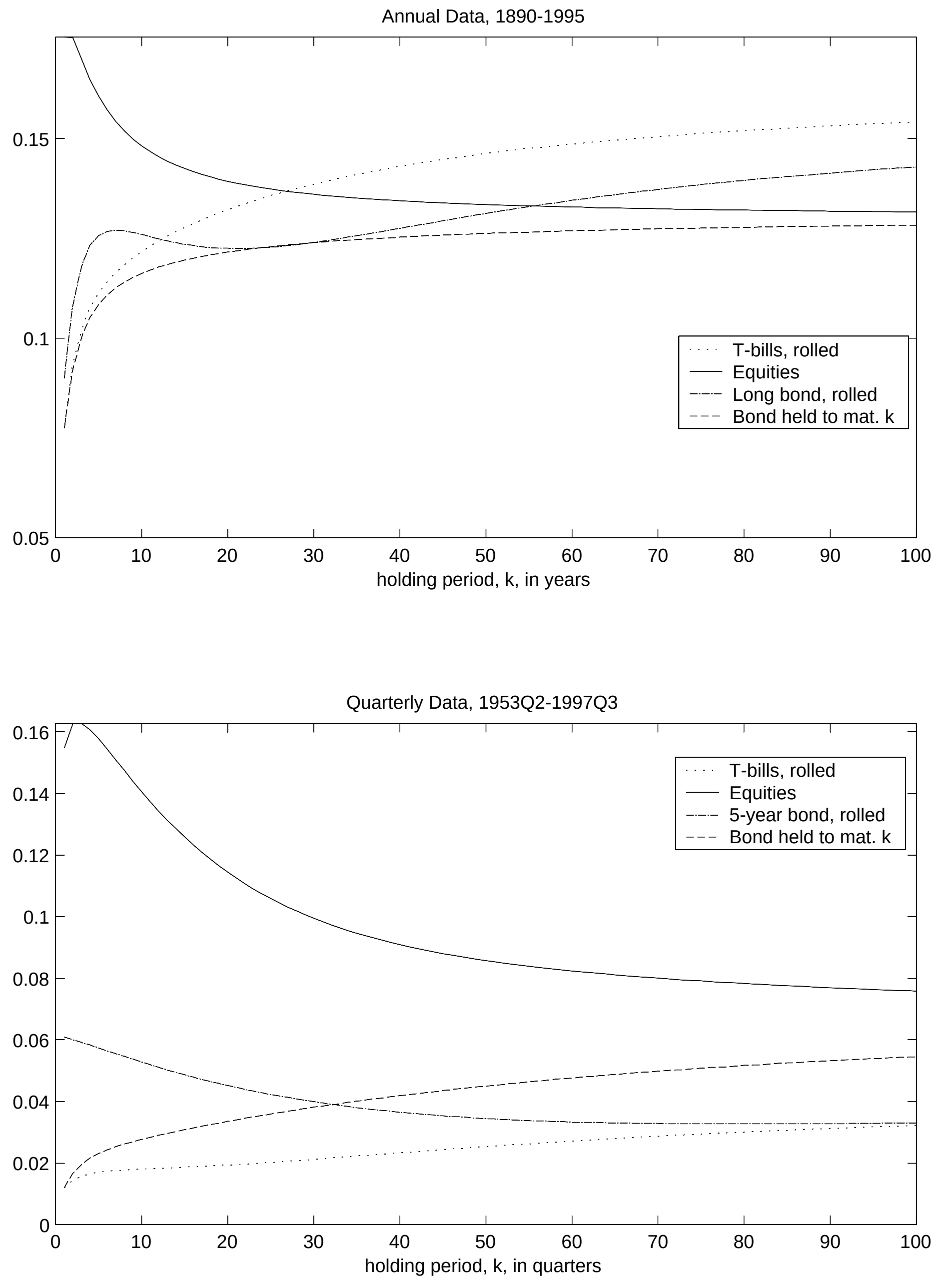

Figure 4.2: Variability of Multiperiod Asset Returns

Siegel (1998). In the quarterly system, stocks are the riskiest asset at all horizons, but their relative risk declines sharply with the horizon. Of the two bond strategies, rolling bonds is riskier at short horizons, but buying and holding is riskier at long horizons since it exposes investors to the persistent variation in inflation that has been characteristic of the postwar period.

These results suggest that long-horizon investors may have a perspective on risk that is very different from the perspective of myopic investors. We explore this issue in the sections that follow.

4.2.3 Strategic allocations to stocks, bonds, and bills

We have shown in section 4.1.2 that the optimal portfolio rule is linear in the vector of state variables. Thus the optimal portfolio allocation to stocks, bonds and bills changes over time. One way to characterize this rule is to examine its mean and volatility. To analyze level effects we compute the mean allocation to each asset as well as the mean hedging portfolio demand for different specifications of the vector of state variables. Specifically, we estimate a series of restricted VAR systems, in which the number of explanatory variables increases sequentially, and use them to calculate mean optimal portfolios for $\psi = 1$ or $1/\gamma$, $\delta = 0.92$ at an annual frequency, and $\gamma = 1$, 2, 5 or 20.

The first VAR system only has a constant term in each regression, corresponding to the case in which risk premia are constant and realized returns on all assets, including the short-term real interest rate, are i.i.d. The second system includes an intercept term, the ex-post real bill rate and log excess returns on stocks and bonds. We then add sequentially the nominal bill rate, the dividend yield and the yield spread. Thus we estimate five VAR systems in total.

Table 4.4 reports the results of this experiment for the annual dataset, for values of the coefficient of relative risk aversion γ equal to 1, 2, 5 and 20, with the intertemporal elasticity of substitution $\psi = 1$. Table 4.5 repeats the results for the quarterly dataset. The entries in each column are mean portfolio demands in percentage points when the explanatory variables in the VAR system include the state variable in the column heading and those to the left of it. For instance, the "constant" column reports mean portfolio allocations when the explanatory variables include only a constant term, that is, when investment opportunities are constant. The right-hand "spread" column gives the case where all state variables are included in the VAR.

Tables 4.4 and 4.5 report results only for selected values of risk aver-

Table 4.4: Mean Asset Demands
Annual Sample (1890 - 1995)

	State Variables:	Constant	AR_t	y_t	$(d-p)_t$	spr_t
$\gamma = 1, \psi = 1, \rho = 0.92$						
Stock	Total Demand	188.44	187.46	189.20	199.43	201.97
	Hedging Demand	0.00	0.00	0.00	0.00	0.00
Bond	Total Demand	127.57	146.72	155.55	155.29	231.73
	Hedging Demand	0.00	0.00	0.00	0.00	0.00
Cash	Total Demand	-216.01	-234.18	-244.75	-254.72	-333.70
	Hedging Demand	0.00	0.00	0.00	0.00	0.00
$\gamma = 2, \psi = 1, \rho = 0.92$						
Stock	Total Demand	98.66	100.81	101.80	132.29	132.47
	Hedging Demand	0.00	2.54	3.02	28.13	27.15
Bond	Total Demand	70.78	89.58	95.75	89.71	53.49
	Hedging Demand	0.00	9.12	12.47	6.58	-64.28
Cash	Total Demand	-69.44	-90.39	-97.55	-122.00	-85.96
	Hedging Demand	0.00	-11.66	-15.49	-34.71	37.13
$\gamma = 5, \psi = 1, \rho = 0.92$						
Stock	Total Demand	44.79	52.42	52.30	74.29	81.38
	Hedging Demand	0.00	7.66	7.79	27.30	34.05
Bond	Total Demand	36.71	53.74	63.75	64.28	-18.87
	Hedging Demand	0.00	13.04	23.82	24.44	-68.29
Cash	Total Demand	18.50	-6.16	-16.05	-38.57	37.49
	Hedging Demand	0.00	-20.70	-31.61	-51.74	34.24
$\gamma = 20, \psi = 1, \rho = 0.92$						
Stock	Total Demand	17.86	29.22	27.93	35.99	40.85
	Hedging Demand	0.00	11.22	10.54	17.59	22.52
Bond	Total Demand	19.67	35.62	48.30	46.95	15.60
	Hedging Demand	0.00	14.80	30.06	28.75	0.40
Cash	Total Demand	62.47	35.15	23.77	17.06	43.55
	Hedging Demand	0.00	-26.02	-40.60	-46.34	-22.92

Table 4.5: Mean Asset Demands
Quarterly Sample (1952Q2 - 1997Q4)

	State Variables:	Constant	AR_t	y_t	$(d-p)_t$	spr_t
$\gamma = 1, \psi = 1, \rho = 0.92^{1/4}$						
Stock	Total Demand	272.65	285.19	289.75	301.76	302.41
	Hedging Demand	0.00	0.00	0.00	0.00	0.00
Bond	Total Demand	42.80	15.90	8.20	2.88	6.54
	Hedging Demand	0.00	0.00	0.00	0.00	0.00
Cash	Total Demand	-215.45	-201.09	-197.95	-204.64	-208.95
	Hedging Demand	0.00	0.00	0.00	0.00	0.00
$\gamma = 2, \psi = 1, \rho = 0.92^{1/4}$						
Stock	Total Demand	136.07	138.76	139.35	313.75	241.40
	Hedging Demand	0.00	-3.63	-5.14	163.32	90.64
Bond	Total Demand	16.28	-36.69	-68.17	-415.78	-465.80
	Hedging Demand	0.00	-39.75	-67.64	-412.63	-464.46
Cash	Total Demand	-52.36	-2.07	28.82	202.03	324.40
	Hedging Demand	0.00	43.37	72.79	249.31	373.82
$\gamma = 5, \psi = 1, \rho = 0.92^{1/4}$						
Stock	Total Demand	54.13	53.07	50.50	578.45	566.02
	Hedging Demand	0.00	-3.63	-6.84	518.86	506.30
Bond	Total Demand	0.38	-30.52	-30.40	-677.15	-1090.92
	Hedging Demand	0.00	-25.87	-24.64	-670.39	-1084.90
Cash	Total Demand	45.49	77.45	79.90	198.69	624.89
	Hedging Demand	0.00	29.51	31.48	151.53	578.60
$\gamma = 20, \psi = 1, \rho = 0.92^{1/4}$						
Stock	Total Demand	13.16	11.49	9.34	358.21	502.51
	Hedging Demand	0.00	-2.37	-4.42	344.09	488.00
Bond	Total Demand	-7.58	-17.53	7.91	-369.26	-799.10
	Hedging Demand	0.00	-9.02	16.28	-360.74	-789.46
Cash	Total Demand	94.42	106.03	82.75	111.05	396.59
	Hedging Demand	0.00	11.40	-11.86	16.66	301.45

sion, but we have also computed portfolio allocations for a continuum of values of risk aversion; Figure 4.3 plots these allocations and their myopic component using the annual VAR with all state variables included. In this figure the horizontal axis shows risk tolerance $1/\gamma$ rather than risk aversion γ, both in order to display the behavior of highly conservative investors more compactly, and because myopic portfolio demands are linear in risk tolerance.

Tables 4.4 and 4.5 enable us to analyze two effects on the level of portfolio demands. By comparing numbers within any column, we can study how total asset allocation and intertemporal hedging demand vary with risk aversion. By comparing numbers within any row, we can examine the incremental effects of the state variables on asset allocation. Here we explore the first topic and leave the second for the next section. To simplify the discussion we focus only on the allocations implied by the full VAR, shown in the right-hand column of the table.

The first set of numbers in Tables 4.4 and 4.5 reports the mean portfolio allocation to stocks, bonds and bills of a logarithmic investor. For this investor, optimal asset allocation is myopic and depends only on the inverse of the variance-covariance matrix of unexpected excess returns and the mean excess return on stocks and bonds. This myopic allocation is long in stocks and bonds in both the annual dataset and the quarterly dataset. However, the ratio of stocks to bonds is close to one in the annual dataset and is about 50 in the quarterly dataset. The preference for stocks in the quarterly dataset is primarily due to the estimated large positive correlation between unexpected excess returns on stocks and bonds in the quarterly dataset. This shifts the optimal myopic allocation towards stocks—the asset with the largest Sharpe ratio. In the annual dataset the correlation between excess bond and stock returns is very low, implying that the optimal portfolio allocation to one asset is essentially independent of the optimal allocation to the other.

Conservative investors, with risk aversion $\gamma > 1$, have an intertemporal hedging demand for stocks. This demand is most easily understood by looking at Figure 4.3, which is based on the annual dataset. In Figure 4.3, the total demand for stocks is a nonlinear, hump-shaped function of risk tolerance $1/\gamma$, while the myopic portfolio demand is, as always, a linear function of $1/\gamma$. Moreover, total stock demand is always larger than myopic portfolio demand for all $1/\gamma < 1$. This implies that intertemporal hedging demand must be a positive, nonlinear function of $1/\gamma$. We can verify this by looking at the hedging demands reported in Tables 4.4 and 4.5. In both datasets, the hedging demand for stocks is always positive and exhibits a

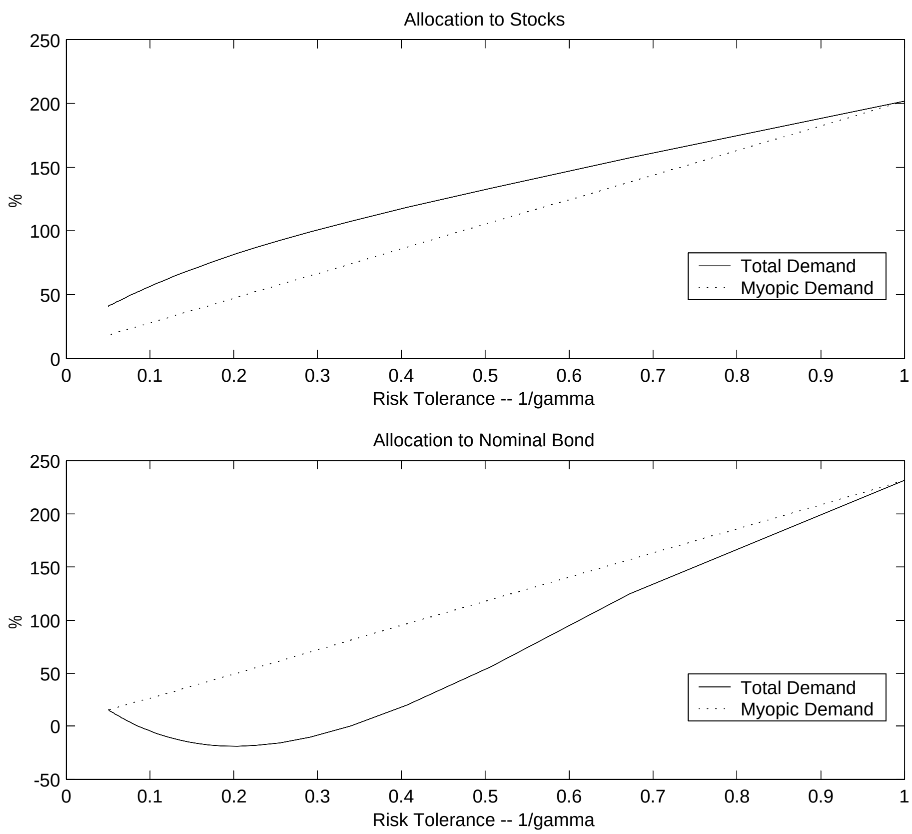

Figure 4.3: Optimal Allocations to Stocks and Nominal Bonds

hump-shaped pattern as a function of $1/\gamma$. These patterns reflect the mean-reversion of stock returns illustrated in Figure 4.2, which is captured in our VAR model by the predictability of stock returns from the dividend-price ratio.

The hedging demand for stock is particularly large in the quarterly dataset (Table 4.5). In fact, in that dataset the hump-shaped hedging demand dominates the linear myopic demand so that total stock demand actually rises with risk aversion for intermediate levels of risk aversion. These results contrast with those of Campbell and Viceira (1999), which are closer to the results reported here for long-term annual data. The main reason for this contrast is that Campbell and Viceira (1999) made an error in estimating their model, understating the predictability of postwar quarterly stock returns and the absolute value of the correlation between innovations to stock returns and dividend yields. This error is explained in Campbell and Viceira (2000), which reports complete corrected results. In addition, the availability of nominal bonds—which are positively correlated with stocks in the quarterly dataset—tends to strengthen hedging demands by allowing investors to offset their equity risks with short positions in bonds.

Intertemporal hedging demands are just as striking for nominal bonds. Table 4.4 shows that the portfolio hedging demand for bonds is negative and exhibits a U-shaped pattern across coefficients of relative risk aversion that eventually reverts to zero. Of course, this shape is also reflected in the total demand for nominal bonds, which is plotted in Figure 4.2. This pattern can be explained by the mean-aversion of bond returns illustrated in Figure 4.2. At a deeper level, it results from the effect of the yield spread on intertemporal hedging demand. We defer this discussion until the next section, where we analyze the effects of individual state variables on portfolio demands.

We have shown allocation results only for the case $\psi = 1$. We have already noted that optimal portfolio demands do not depend on the elasticity of intertemporal substitution except through the loglinearization parameter ρ. Results not reported here for the power utility case show that this indirect effect is quantitatively insignificant. The allocations for power utility are almost indistinguishable from those in Tables 4.4 and 4.5.

We can also use our model to decompose the variability of asset demands. We can write the optimal portfolio rule as

$$\alpha_{it} = \alpha_{it}^{m} + \alpha_{it}^{h}, \tag{4.20}$$

where i denotes stocks or bonds, m denotes myopic demand, and h denotes hedging demand. Thus,

$$\operatorname{Var}(\alpha_{it}) = \operatorname{Var}(\alpha_{it}^m) + \operatorname{Var}\left(\alpha_{it}^h\right) + 2\operatorname{Cov}\left(\alpha_{it}^m, \alpha_{it}^h\right). \tag{4.21}$$

We have calculated this variance decomposition for the case $\gamma = 5$ and $\psi = 1$. The hedging component explains at most 23% of the total variation in portfolio demand for both stocks and bonds in the annual dataset, and 14% in the quarterly dataset. Thus hedging portfolio demand is much more stable than total portfolio demand. Kim and Omberg (1996) and Campbell and Viceira (1999) give an intuitive explanation for this result, showing that hedging demand can change sign only in extreme circumstances where investors have replaced their normal long positions with short positions in risky assets. To a first approximation, intertemporal hedging shifts the intercept of risky asset demand rather than the slope with respect to state variables; put another way, long-term investors should "market-time" just as aggressively as short-term investors.

Which state variables matter?

The analysis so far has focused on the shape of asset demands and their hedging components. It is equally important to understand the effects of various state variables on the level and variability of asset demands. To analyze the level effects of state variables, we can compare average portfolio demands across rows in Tables 4.4 and 4.5.

Table 4.4 shows that there are important changes in the magnitude of hedging demands as we consider new state variables in the investor information set. In the case of stocks, hedging demand is very small when only lagged Treasury bill rates (either real or nominal) and excess returns on bonds and stocks are included in the VAR. It shoots up dramatically when the dividend-price ratio is introduced into the VAR as a regressor. The inclusion of the yield spread has mixed effects in the annual dataset, and negative effects in the quarterly dataset.

The correlation structure shown in Table 4.4 helps explain these results. In the full annual VAR system, there is a strong negative correlation between unexpected excess returns on stocks and shocks to the dividend-price ratio, while the magnitude of all other correlations in the table is much smaller. These correlations are not sensitive to the inclusion or exclusion of state variables in the VAR. The presence of the dividend-price ratio in the investor information set increases the hedging demand for stocks because negative shocks to the dividend-price ratio, which drive down expected returns on stocks, tend to coincide with positive realized excess returns on stocks. This negative correlation is even stronger in the quarterly dataset, which makes the pattern for hedging demands more pronounced in this dataset.

In the case of bonds, the yield spread has a tremendous negative impact on hedging demand. In fact, it changes the sign of hedging demand from positive to negative. Table 4.4 again explains this result. In both datasets, the yield spread is the most important forecasting variable for bond returns, and its innovations are positively correlated with excess bond returns. This correlation produces a negative hedging demand for bonds, since negative shocks to expected future bond returns tend to coincide with negative current bond returns. The magnitude of the correlation is larger in the quarterly dataset, which is why hedging demand is more negative in this dataset.[6]

Hedging demands can also be understood by reference to Figure 4.2, which shows that stock returns are mean-reverting, while nominal bond returns are mean-averting. A univariate representation of excess stock returns will have a negative correlation between expected and unexpected excess returns on stocks, while a univariate representation of excess bond returns will have a positive correlation between expected and unexpected excess returns on bonds. This makes stocks an attractive asset for conservative investors who seek to hedge intertemporally, while it makes nominal bonds a fundamentally unattractive asset.

We can also analyze the importance of each state variable for the variability of asset demands. In the case $\gamma = 5$ and $\psi = 1$, the dividend-price ratio explains 80% of the variance of total demand for stocks in the annual sample, and 91% in the quarterly sample. The dividend-price ratio plays a much less important role in explaining the variability of the portfolio demand for bonds, which is driven primarily by the lagged excess stock return and the yield spread.

In summary, our results indicate that the most important state variable determining the mean and volatility of stock demand is the dividend yield, while the yield spread is more important for bonds. The dividend yield generates a large positive intertemporal hedging demand for stocks, while

[6]The positive correlation of bond and stock returns in the quarterly dataset also means that positive hedging demand for stocks tends to produce negative hedging demand for bonds in that dataset. Note also that shocks to nominal bill yields are highly negatively correlated with unexpected excess bond returns in both samples, but the coefficient on the nominal bill rate in the excess bond return equation in the VAR is small and not statistically significant. This lack of predictive power means that the inclusion of the nominal bill rate has a relatively small effect on the hedging demand for bonds.

the yield spread generates a large negative intertemporal hedging demand for bonds. Aït-Sahalia and Brandt (2000) also find that these variables are important determinants of optimal portfolio choice, though they find that the role of the dividend yield weakens if the late 1990's are included in the sample.

4.2.4 Strategic asset allocation with inflation-indexed bonds

Our results so far imply that the intertemporal hedging demand for long-term bonds is negative. This contrasts with the conventional investment advice that conservative long-term investors should hold bonds to obtain a stable stream of income, disregarding short-run fluctuations in capital value. There are two possible reasons for the discrepancy between our results and conventional wisdom. First, the conventional wisdom disregards the distinction between nominal and inflation-indexed bonds. In the presence of significant inflation risk, long-term nominal bonds are not suitable assets for conservative long-term investors as we showed in Chapter 3. Second, the model we use in this chapter has a general dynamic structure in which either stocks or bonds might be good hedges for predictable variation in stock and bond returns. Conventional investment advice may be based on the presumption that bonds are the best hedges for predictable variation in returns on all risky assets; the model of Chapter 3 explicitly assumes this.

To determine which of these explanations is correct, we now extend our model to include an inflation-indexed perpetuity in the menu of available assets. This requires us to construct hypothetical real bond returns, because indexed bonds have only been recently issued by the US Treasury and thus data on indexed bonds are very limited. The VAR framework is well suited for this purpose, provided that we make the assumption that expected real returns on real bonds of all maturities and the expected real return on short-term nominal bills differ only by a constant. This amounts to assuming that the inflation risk premium on nominal bills is constant. We now briefly describe the construction procedure, which is adapted from the work of Campbell and Shiller (1996) and described in detail in the Appendix.

We first use the estimates of the coefficient matrices in the VAR to construct returns on hypothetical real perpetuities according to the procedure outlined in the Appendix. The procedure assumes a zero inflation risk premium. As noted in Campbell and Shiller (1996), if the inflation risk premium is not zero but constant, the procedure will miss the average level of the yield curve, but will still capture the dynamics of the curve. This is important, because intertemporal hedging demand depends sensitively on the dynamics of asset returns. With the correct dynamics in hand, we adjust the mean return by setting the Sharpe ratio of the real consol bond to the Sharpe ratio of nominal bonds.[7] Finally, we include the imputed excess return on real perpetuities in a new VAR system that includes both nominal bonds and real consols.

Table 4.6 reports the resulting mean asset demands for values of γ equal to 1, 2, 5, 20 and 2000 and $\psi = 1$. We include the case $\gamma = 2000$ because we want to study asset demand for infinitely risk averse investors, which we proxy using this large value of γ. We also report mean asset allocations under constant investment opportunities. To simplify the discussion and to save space, we include results only for the annual dataset. Figure 8 plots the allocations implied by the full VAR for a continuum of values of γ.

We start by looking at the optimal portfolio of a myopic logarithmic investor. This investor should hold a short position in the inflation-indexed perpetuity, despite the fact that the mean excess return on this asset is positive by construction. This allocation is the result of a large, positive correlation between excess returns on stocks and excess returns on the real perpetuity (shown in Appendix D), which makes it optimal for a logarithmic investor to short the real perpetuity to increase her allocation to stocks, the asset with the largest Sharpe ratio.

We can learn about the myopic allocations of non-logarithmic investors by looking at the allocations under constant investment opportunities shown in the "constant" column. Investors with $\gamma > 1$ have a myopic demand for real perpetuities that is not proportional to the optimal allocation of the logarithmic investor. In fact, it even changes sign and becomes positive for moderately risk averse investors. This is driven by the fact that the short-term bill is risky in real terms, so the portfolio with the smallest short-term risk is a combination, with roughly equal positive weights, of the short-term bill and the real perpetuity.

The "spread" column in Table 4.6 shows total portfolio demands with time-varying investment opportunities. The total portfolio demand for the real consol bond is increasing in risk aversion, approaching 100% of the portfolio as the investor becomes infinitely conservative. By contrast, the total portfolio demand for stocks, the nominal bill and the nominal bond are decreasing in γ, approaching 0% as the investor becomes infinitely conservative. Thus inflation-indexed bonds drive out cash from the portfolios

[7]We have also considered setting the Sharpe ratio of the real consol bond equal to zero and setting it equal to the Sharpe ratio of stocks. These choices affect myopic asset demands, but do not have noticeable effects on intertemporal hedging demands. Results are available from the authors upon request.

Table 4.6: Mean Asset Demands with Hypothetical Real Bonds

State Variables:	Constant	spr_t
$\gamma = 1, \psi = 1, \rho = 0.92$		
Stocks	188.49	221.00
Real Consol Bond	-76.96	-103.51
Nominal Bond	144.71	290.94
Cash	-156.25	-308.43
$\gamma = 2, \psi = 1, \rho = 0.92$		
Stocks	94.05	134.25
Real Consol Bond	-11.63	-21.99
Nominal Bond	70.13	73.28
Cash	-52.54	-85.54
$\gamma = 5, \psi = 1, \rho = 0.92$		
Stocks	37.38	70.75
Real Consol Bond	27.57	46.71
Nominal Bond	25.38	-36.43
Cash	9.68	18.97
$\gamma = 20, \psi = 1, \rho = 0.92$		
Stocks	9.04	21.95
Real Consol Bond	47.17	84.40
Nominal Bond	3.00	-20.03
Cash	40.79	13.68
$\gamma = 2000, \psi = 1, \rho = 0.92$		
Stocks	-0.31	0.09
Real Consol Bond	53.63	96.92
Nominal Bond	-4.38	6.61
Cash	51.05	-3.62

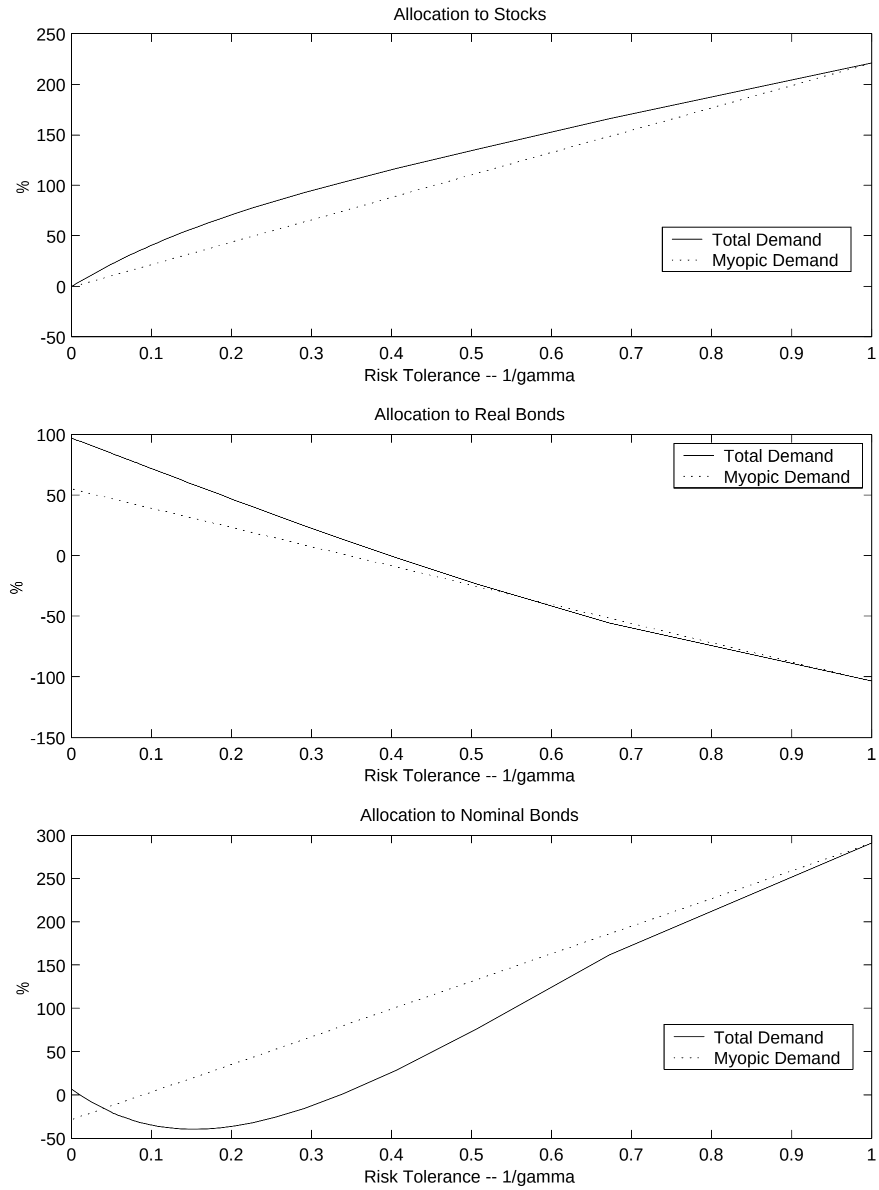

Figure 4.4: Optimal Allocations to Stocks, Real Consols, and Nominal Bonds

of conservative investors. This is a generalization of the result of Chapter 3 to a world in which both interest rates and expected excess returns are time-varying.[8]

4.3 Conclusion

In this chapter we have explored optimal investment strategies when both riskless interest rates and risk premia change over time. In this situation a long-term investor with constant risk aversion should both take advantage of and hedge against variations in investment opportunities. In the presence of mean-reverting stock returns, the strategic equity allocation is higher on average than the optimal myopic equity allocation, and it responds slightly more strongly to changes in the equity premium; thus it involves an element of market timing.

It is interesting to relate these results to recent discussions of stock market risk. Equities have traditionally been regarded as risky assets. They may be attractive because of their high average returns, but these returns represent compensation for risk; thus equities should be treated with caution by all but the most aggressive investors. In recent years, however, several authors have argued that equities are actually relatively safe assets for investors who are able to hold for the long term. We have already quoted Jeremy Siegel (1994) on this point; a more extreme version of this revisionist view is promoted by James Glassman and Kevin Hassett, who argue in their recent book *Dow 36,000* (1999) that stocks are just as safe as bonds or Treasury bills.

The revisionist view that stocks are safe assets is based on evidence that excess stock returns are less volatile when they are measured over long holding periods. Mathematically, such a reduction in stock market risk at long horizons can only be due to mean-reversion in excess stock returns, which is equivalent to time-variation in the equity premium. Yet revisionist investment advice typically ignores the implications of a time-varying equity premium. Siegel (1994) recommends an aggressive buy-and-hold strategy, like the horizontal line in Figure 4.1 but shifted upwards to reflect the reduced risk of stocks for long-term investors. The optimal policy is instead

[8]Note that the allocation to the real consol bond for an investor with an extremely large coefficient of relative risk aversion does not equal 100% exactly. This is due primarily to the fact that the investor we consider in Table 5 has unit, not zero, elasticity of intertemporal substitution. There is also a small effect caused by the fact that the VAR system in Table 5 does not exactly capture the information set we used to construct the long-term real bond yield.

the sloped line marked "Strategic investor" in Figure 4.1.

The difference between the optimal strategy and the strategy recommended by Siegel is particularly dramatic at times like the present, when recent stock returns have been spectacular. At such a time, the optimal equity allocation may be no higher—it may even be lower—than the allocation implied by a traditional short-term portfolio analysis. To put it another way, investors who are attracted to the stock market by the prospect of high returns combined with low long-term risk are trying to have their cake and eat it too. If expected stock returns are constant over time, then one can hope to earn high stock returns in the future similar to the high returns of the past; but in this case stocks are much riskier than bonds in the long term, just as they are in the short term. If instead stocks mean-revert, then they are relatively safe assets for long-term investors; but in this case future returns are likely to be meagre as mean-reversion unwinds the spectacular stock market runup of the past decade.

It is important to keep in mind two limitations of our analysis in this chapter. First, we ignore portfolio constraints that might prevent investors from short-selling or from borrowing to invest in risky assets. The Siegel strategy of buying and holding stocks might be much closer to optimal for an aggressive investor who cannot borrow to leverage a stock market position, and who therefore normally holds the maximum 100% weight in equities.

Second, we have studied a partial equilibrium model. We have solved the microeconomic problem of an investor facing exogenous asset returns, but we do not show how these asset returns could be consistent with general equilibrium. The difficulty is particularly severe in this chapter, since we find that all investors should change their portfolio allocations in the same direction as state variables change, regardless of their preferences. That is, all investors should buy and sell assets at the same time. This cannot be consistent with a general equilibrium model that makes realistic assumptions about asset supplies.

One possible resolution of this difficulty is that the representative investor has different preferences from those assumed here, perhaps the habit-formation preferences of Campbell and Cochrane (1999) that can generate shifts in risk aversion and hence changing risk premia with a constant riskless interest rate. Under this interpretation the results of this chapter should be used only by investors with constant risk aversion, who cannot be typical of the market as a whole.

Chapter 5

Strategic Asset Allocation in Continuous Time

In the first part of this book we have developed a discrete-time model that can be used to analyze optimal long-term portfolio choice when the conditions for myopic portfolio choice fail. In particular, we have asked how long-term investors should react to time variation in interest rates and risk premia. This chapter extends the previous analysis in two ways.

First, we show how to approach similar problems in a continuous-time framework. The use of continuous-time mathematics to analyze dynamic portfolio choice has a long tradition that goes back at least to the seminal work by Robert Merton (1969, 1971, 1973). Duffie (1996) and Merton (1990) provide a general treatment of portfolio choice in continuous time. We show that, when exact analytical solutions are not available, we can still obtain approximate analytical solutions of the same nature as the ones we have presented in previous chapters. Furthermore, when investors' preferences are characterized by a recursive utility function, we can obtain exact analytical solutions for investors with unit elasticity of intertemporal substitution.

Second, we explore the investment implications of time-varying risk. A continuous-time framework is convenient for this purpose because continuous-time models of stochastic volatility are parsimonious and readily restrict volatility to be positive. Our analysis of this problem closely follows Chacko and Viceira (1999). We continue to assume that investors have financial wealth but no uninsurable labor income risk.

This chapter is written at a higher technical level than other chapters in the book. Readers who are unfamiliar with continuous-time mathematics—in particular, Itô's Lemma—should consult a primer such as Neftci (1996) or should skip this chapter altogether.

The organization of the chapter is as follows. Section 5.1 develops the dynamic programming approach to optimal consumption and portfolio choice in continuous time, as introduced originally by Merton. Section 5.1.1 explains the Bellman optimality principle in continuous time, section 5.1.2 introduces an example that is a continuous-time equivalent of the real term-structure model in Chapter 3, and section 5.1.3 derives a continuous-time loglinear approximate solution to the model, a continuous-time equivalent of the solution derived in Chapter 3.

Section 5.2 explains the Cox-Huang solution method, the leading continuous-time alternative to the dynamic programming approach. Section 5.2.1 explains the role of the stochastic discount factor (SDF) in continuous-time models, section 5.2.2 shows how the properties of the SDF help one to solve dynamic portfolio choice problems, and section 5.2.3 revisits the term-structure example of section 5.1.2.

For simplicity, both sections 5.1 and 5.2 work with time-separable power utility. Section 5.3 presents recursive utility, the continuous-time equivalent of the Epstein-Zin preferences introduced in Chapter 2. Finally, section 5.4 applies the methods of this chapter to solve a long-term portfolio choice problem with time-varying stock market volatility.

5.1 The Dynamic Programming Approach

5.1.1 The Bellman Optimality Principle

We start by deriving the Bellman equation for optimality in a simple setting. The Bellman optimality principle is a useful tool for solving dynamic portfolio problems, because it allows the transformation of a dynamic optimization problem into a differential equation, for which several solution methods are available.

For notational convenience, we assume there are only two assets available to the investor, a risky asset with instantaneous total return dP_t/P_t[1] and an instantaneously riskless asset with return dB_t/B_t. There is also a single state variable S_t driving the dynamics of the investment opportunity set. It is conceptually straightforward to extend the analysis to multiple assets and state variables.

[1] If the risky asset does not pay dividends, P_t is simply the price of the asset. If it does pay dividends, P_t represents an index whose instantaneous rate of change dP_t/P_t equals the instantaneous total return on the asset.

We assume that both returns and the state variable follow diffusion processes:

$$\frac{dP_t}{P_t} = \mu_P(S,t)\,dt + \sigma_P(S,t)\,dZ_{P,t}, \tag{5.1}$$

$$\frac{dB_t}{B_t} = r(S,t)\,dt, \tag{5.2}$$

$$dS_t = \mu_S(S,t)\,dt + \sigma_S(S,t)\,dZ_{S,t}, \tag{5.3}$$

with $dZ_{P,t}dZ_{S,t} = \rho_{PS}(S,t)\,dt$. Note that the drifts, volatilities, and correlations of these processes may be functions of the state variable and time. In the following equations, however, we often omit this dependence, writing for example μ_P instead of $\mu_P(S,t)$; this simpler notation is less careful but more readable.

Given time-separable preferences defined over consumption, and initial wealth $W_0 > 0$, we can formulate the optimal portfolio and consumption problem for a long-term investor as

$$\max_{C,\alpha} \mathrm{E}_0\left[\int_0^\infty U(C,t)\,dt\right] \tag{5.4}$$

subject to the continuous-time intertemporal budget constraint

$$dW = \left[\left(\alpha\left(\mu_P - r\right) + r\right)W - C\right]dt + \alpha W \sigma_P dZ_P \tag{5.5}$$

and the constraints $W_0 > 0$, $W_t > 0$, and $C_t > 0$. Here α denotes the fraction of wealth invested in the risky asset, and C denotes consumption.

Let $J(W,S,t)$ denote the maximized utility function, or value function, of this problem. Merton (1971, 1973) shows that the Bellman principle implies:

$$0 = \max_{\alpha,C}\left\{U(C,t) + \frac{1}{dt}\mathrm{E}_t\left[dJ(W,S,t)\right]\right\}. \tag{5.6}$$

At the optimum, the investor has perfectly traded off the value of present and future consumption. Consumption today is achieved at the expense of current resources that otherwise could increase consumption in the future. The investor chooses a level of current consumption whose utility value exactly offsets the expected utility cost of lost future consumption.

Under suitable regularity conditions (Merton 1990), Itô's Lemma implies that

$$\begin{aligned} dJ(W,S,t) &= J_W dW + J_S dS + (\partial J/\partial t)dt \\ &\quad + \frac{1}{2}J_{WW}(dW)^2 + J_{WS}dW dS + \frac{1}{2}J_{SS}(dS)^2. \end{aligned} \tag{5.7}$$

Here we use subscripts to denote partial derivatives, except that we write out the partial time derivative of J explicitly as $\partial J/\partial t$ to avoid any possibility of confusion with the value of the function J at time t. Using the stochastic differential equations for dS and dW given in (5.3) and (5.5), and the rules of stochastic calculus, we can easily compute an expression for the expected instantaneous change in the value function. Substitution of the expression for $\mathrm{E}_t[dJ(W,S,t)]/dt$ into (5.6) gives an equation that depends on C, α, and the value function:

$$\begin{aligned} 0 &= \max_{\alpha,C}\left\{U\left(C,t\right)+J_W\left[\left(\alpha\left(\mu_P-r\right)+r\right)W-C\right]+J_S\mu_S+\partial J/\partial t\right. \\ &\left.+\frac{1}{2}J_{WW}\alpha^2W^2\sigma_P^2+J_{WS}\alpha W\sigma_P\sigma_S\rho_{PS}+\frac{1}{2}J_{SS}\sigma_S^2\right\}. \end{aligned} \tag{5.8}$$

Merton (1969) notes that this equation must verify the boundary condition

$$\lim_{t\to\infty}\mathrm{E}_0\left[J\left(W,S,t\right)\right]=0. \tag{5.9}$$

This is a condition for the convergence of the integral in (5.4). It is a transversality condition ensuring that the value function is bounded in the limit, i.e., that there is no investment strategy that allows the investor to achieve infinite utility.

We can compute the first-order conditions of the problem by taking derivatives of (5.8) with respect to C and α. We obtain a pair of expressions for consumption and portfolio choice as a function of wealth:

$$U_C = J_W \Rightarrow C = U_C^{-1}\left(J_W\right), \tag{5.10}$$

$$\begin{aligned} \alpha &= \frac{1}{-J_{WW}W/J_W}\left(\frac{\mu_p-r}{\sigma_P^2}\right) \\ &-\frac{J_{WS}}{J_{WW}W}\left(\frac{\sigma_S}{\sigma_P}\rho_{PS}\right). \end{aligned} \tag{5.11}$$

Here consumption and the derivatives of the value function all depend on the variables W, S, and t, while the riskless rate and the moments of the risky asset return all depend on S and t. For simplicity these dependences are omitted from the notation, but they should not be forgotten.

Equation (5.10) determines the optimal consumption policy. It is known as the "envelope condition," because it implies that at the optimum an extra unit of current consumption is as valuable to the investors as an extra unit of wealth to finance future consumption.

Equation (5.11) determines the optimal portfolio allocation to the risky asset. It is the continuous-time counterpart of (3.14). The first term of this equation is the familiar myopic portfolio rule: The optimal allocation to the risky asset is proportional to the asset's risk premium, and inversely proportional to its volatility and the relative risk aversion of the investor's value function. The second term is the intertemporal hedging component. It is non-zero as long as investment opportunities are time-varying ($\sigma_S > 0$), they are correlated with instantaneous realized returns on the risky asset ($\rho_{PS} \neq 0$), and they affect the marginal utility of wealth ($J_{WS} \neq 0$).

Equations (5.10) and (5.11) are not a complete solution to the model because they depend on the value function, which is still unknown. However, substitution of these expressions for optimal consumption and portfolio choice back into the Bellman equation (5.8) delivers a second-order partial differential equation (PDE) for the value function $J(W, S, t)$. Once we have solved this equation for the value function, we can obtain the optimal policies by substituting the value function into the first-order conditions for consumption and portfolio choice.

Unfortunately, it is not generally straightforward to find an analytical solution for the PDE that gives the value function. This type of equation is solved analytically using the method of undetermined coefficients. That is, one makes a guess about the functional form of the solution, and shows that it verifies the partial differential equation for some values of the parameters defining the functional guess. However, finding such a function is not always easy. In some cases, it is possible to transform the PDE into an ordinary differential equation (ODE), and there are handbooks such as Polyanin and Zaitsev (1995) that help identify ordinary differential equations with known exact solutions. But if no exact solution is known, it is necessary to resort to numerical algorithms such as those explained in Judd (1998) or Rogers and Talay (1997).

The special cases with known analytical solutions can be listed quite briefly. Merton (1969, 1971, 1973) showed that equation (5.6) has an exact analytical solution if investors' utility of consumption is logarithmic, or if utility is power and investment opportunities are constant. In both cases, as we noted in Chapter 2, the intertemporal asset allocation problem is essentially equivalent to a one-period problem, and optimal portfolio choice is myopic.

More recently, Kim and Omberg (1996) and Liu (1998) have shown that we can solve the problem exactly provided that markets are complete—which in our simplified model, with only one state variable, requires perfect correlation between the state variable and the stock return—and investors maximize the utility of terminal wealth at some fixed horizon. Using the method of Cox and Huang (1989), Wachter (1999) has extended this solution to the case of investors who maximize utility over consumption each period. Schroder and Skiadas (1997) and Fisher and Gilles (1998) also explore the implications of complete markets for optimal portfolio choice when investors have recursive utility. Chacko and Viceira (1999) show that we can obtain an exact solution to the continuous-time problem without assuming that markets are complete, provided that investors' elasticity of intertemporal substitution is one.

In all other cases, there are no known exact analytical solutions. However, we can resort to an approximate analytical solution method of the sort we introduced in previous chapters for discrete-time consumption and portfolio choice problems. We illustrate this approach in the next section by studying an example, the continuous-time counterpart of the model with time-varying real interest rates introduced in Chapter 3.

5.1.2 Consumption and portfolio choice with time-varying interest rates and power utility.

The derivation of the Bellman equation shown in the previous section is general, except for the limited number of assets and state variables. To apply this solution method to a particular problem we need to be more specific about the direct utility function $U(.)$ and the investment opportunity set. In this section we solve a simple example, where investors have time-separable power utility over consumption with constant relative risk aversion γ, $U(C_t) = C_t^{1-\gamma}/(1-\gamma)$, and where time-variation in investment opportunities is created by a time-varying real interest rate, so that $S_t = r_t$. To keep our notation as simple as possible, we also assume that the investor can choose only between an instantaneous short-term real bond paying rdt and a long-term real bond.

We assume that the real interest rate follows an Ornstein-Uhlenbeck process as in Vasicek (1977):

$$dr_t = \kappa_r \left(\theta_r - r_t\right) dt + \sigma_r dZ_r, \tag{5.12}$$

where $\kappa_r \in (0,1)$ and $\theta_r > 0$. Let $P(r_t, T-t)$ denote the price of a real zero-coupon bond with $T-t$ periods to maturity. Vasicek (1977) shows that no-arbitrage implies the following dynamics for bond returns:

$$\frac{dP\left(r_t, T-t\right)}{P\left(r_t, T-t\right)} = \left[r_t + \lambda b\left(T-t\right)\right] dt - b\left(T-t\right) \sigma_r dZ_r, \tag{5.13}$$

where λ determines the risk premium on the bond and $b(T-t) = \kappa_r^{-1}(1 - \exp\{-\kappa_r(T-t)\}) \equiv -P_r/P$. Thus real bonds in this model have a constant expected excess return equal to $\lambda b(T-t)$ and a constant instantaneous return volatility equal to $b(T-t)\sigma_r$. There is only one source of uncertainty in this model, and bond returns are perfectly negatively correlated with the short-term real interest rate. This is the continuous-time counterpart of the real-interest-rate model in Chapter 3.

The investor's dynamic portfolio and consumption problem can now be written as

$$\max_{C,\alpha} \mathrm{E}_0 \left[\int_0^\infty \mathrm{e}^{-\beta t} \frac{C_t^{1-\gamma}}{1-\gamma} dt \right] \tag{5.14}$$

subject to

$$dW_t = \left[\alpha\lambda b \left(T-t\right) + r_t W_t - C_t \right] dt - \alpha W_t b \left(T-t\right) \sigma_r dZ_r. \tag{5.15}$$

To find the optimal consumption and portfolio policies for this model we follow the steps described in Section (5.1.1). The Bellman equation for this problem is

$$\begin{aligned} 0 \;=\; & \max_{\alpha,C} \left\{ \mathrm{e}^{-\beta t} \frac{C^{1-\gamma}}{1-\gamma} + J_W \left(\alpha\lambda b W + rW - C\right) + J_r \kappa_r \left(\theta_r - r\right) + \partial J/\partial t \right. \\ & \left. + \frac{1}{2} J_{WW} \alpha^2 W^2 b^2 \sigma_r^2 - J_{Wr} \alpha W b \sigma_r^2 + \frac{1}{2} J_{rr} \sigma_r^2 \right\}, \end{aligned} \tag{5.16}$$

where we have set $b(T-t) \equiv b$ for brevity.

From this Bellman equation, the first-order conditions for optimal consumption and portfolio choice are:

$$C = \left(\mathrm{e}^{\beta t} J_W \right)^{-(1/\gamma)} \tag{5.17}$$

$$\alpha = \left(\frac{1}{-J_{WW} W / J_W} \right) \frac{\lambda}{b\sigma_r^2} + \left(\frac{J_{Wr}}{J_{WW} W} \right) \frac{1}{b}. \tag{5.18}$$

Substitution of these conditions into equation (5.16) gives a second-order ordinary differential equation (ODE) for the value function $J(W, r, t)$ shown in the mathematical appendix. To solve this equation we guess that the value function takes the form

$$J\left(W, r, t\right) = \mathrm{e}^{-\beta t} H\left(r_t\right)^\gamma \frac{W_t^{1-\gamma}}{1-\gamma}, \tag{5.19}$$

where $H(r_t)$ is a function only of the instantaneous interest rate. This guess implies, after some simplifications, another ODE for the function $H(r_t)$:

$$\begin{aligned} 0 &= \frac{\gamma}{1-\gamma}H^{-1} + \left(\frac{\lambda}{2\gamma\sigma_r^2} - \frac{\beta}{1-\gamma} + r\right) + \left(\frac{\gamma\kappa_r}{1-\gamma}\left(\theta_r - r\right) - \lambda\right)\left(\frac{H_r}{H}\right) \\ &\quad + \frac{\gamma\sigma_r^2}{2\left(1-\gamma\right)}\left(\frac{H_{rr}}{H}\right). \end{aligned} \tag{5.20}$$

Equation (5.20) is a non-homogeneous ODE, whose associated homogeneous equation belongs to the degenerate hypergeometric equation class. This equation has an exact solution given in Polyanin and Zaitsev (1995). Unfortunately this solution is a complicated expression involving gamma functions which is extremely hard to interpret.

5.1.3 An approximate analytical solution

We now show that it is possible to find an approximate analytical solution to the problem. The solution is based on a log-linear expansion of the consumption-wealth ratio around its unconditional mean. This is exactly the same type of approximation that we used in Chapters 3 and 4, but instead of using it to linearize the intertemporal budget constraint, we use it here to solve the Bellman equation.

To understand the approach, note that the envelope condition (5.17) implies

$$\frac{C_t}{W_t} = \exp\left\{c_t - w_t\right\} = H\left(r_t\right)^{-1}, \tag{5.21}$$

where $c_t - w_t = \log\left(C_t/W_t\right)$. Therefore, we can approximate $H(r_t)^{-1}$ as

$$H\left(r_t\right)^{-1} \approx h_0 + h_1\left(c_t - w_t\right), \tag{5.22}$$

where $h_0 = \exp\{\overline{c-w}\}\,[1 - (\overline{c-w})]$, $h_1 = \exp\{\overline{c-w}\}$, and $(\overline{c-w}) = \mathrm{E}[c_t - w_t]$. Substituting (5.22) for $H(r_t)^{-1}$ in the first term of (5.20), it is easy to see that the resulting ODE has a solution of the form $H(r_t) = \exp\{C_0 + C_1 r_t\}$. This implies that the log consumption-wealth ratio is linear in the riskless real interest rate: $c_t - w_t = -C_1 r_t - C_0$.

Our approach replaces the term that causes the non-linear ODE (5.20) to be non-solvable analytically with a log-linear approximation. Thus we transform the equation into another ODE with a known analytical solution. If the log-linear approximation is accurate, the exact analytical solution to the approximate ODE will also verify the original ODE subject to some approximation error, and can be regarded as an approximate analytical solution to the original ODE. We will also show that the approximation error is zero for the special cases of log utility and constant investment opportunities.

The approximate ODE leads to two algebraic equations for C_1 and C_0 given in the Appendix. The first equation is linear in the coefficient C_1, with solution

$$C_1 = -\left(1 - \frac{1}{\gamma}\right) \frac{1}{h_1 + \kappa_r}. \tag{5.23}$$

The second equation involves both coefficients, but it is linear in C_0 given C_1. Its solution is a function of all the parameters in the model.

The approximate solution implies the value function

$$J(W, r, t) = \exp\{-\beta t + \gamma C_0 + \gamma C_1 r_t\} \frac{W_t^{1-\gamma}}{1-\gamma}, \tag{5.24}$$

and the optimal policies

$$c_t - w_t = -C_0 + \left(1 - \frac{1}{\gamma}\right) \frac{1}{h_1 + \kappa_r} r_t, \tag{5.25}$$

$$\alpha = \frac{1}{\gamma} \frac{\lambda}{b(T-t)\sigma_r^2} + \left(1 - \frac{1}{\gamma}\right) \frac{1}{b(T-t)(h_1 + \kappa_r)}. \tag{5.26}$$

This is the continuous-time equivalent of the discrete-time approximate solution given in Chapter 3. The optimal portfolio policy is a weighted average of two terms, with weights given by the investor's coefficient of relative risk tolerance and one minus this coefficient. It is straightforward to show that the solution is exact in the log utility case where $\gamma = 1$. In this case, $C_t/W_t = \beta$ and $\alpha = \lambda / b(T-t)\sigma_r^2$. This is the exact solution implied by the Bellman equation (5.20) with $\gamma = 1$.

5.2 The Cox-Huang Approach

Cox and Huang (1989) have suggested an alternative approach to intertemporal consumption and portfolio choice that takes advantage of the properties of the stochastic discount factor under complete markets. This approach works by transforming dynamic problems into a static problem whose unknown is optimally invested wealth rather than the value function. This transformation delivers a differential equation for optimally invested wealth that is often easier to solve than the Bellman equation for the value function. In this section we offer a "hands-on" explanation of how this approach works. We start by describing the properties of the stochastic discount factor in continuous time.

5.2.1 The stochastic discount factor in continuous time

The stochastic discount factor in continuous time is defined as the process M_t such that for any security with price V_t and instantaneous payoff X_s we have:

$$V_t = \mathrm{E}_t\left[\frac{M_s}{M_t}X_s\right], \qquad s \geq t. \tag{5.27}$$

The stochastic discount factor is also known as the pricing kernel or state-price density. An important property of the continuous-time stochastic discount factor is that it is unique if markets are complete and there are no opportunities for arbitrage in the economy.

If we are considering a security that does not pay dividends, we have $X_s = V_S$, and (5.27) becomes

$$M_t V_t = \mathrm{E}_t\left[M_s V_s\right], \tag{5.28}$$

which in turn implies that $M_t V_t$ follows a martingale:

$$\mathrm{E}_t\left[d\left(M_t V_t\right)\right] = 0. \tag{5.29}$$

If the security pays an instantaneous dividend of $D_t dt$ each period, we have $X_s = V_s + D_s dt$, and (5.27) becomes

$$\mathrm{E}_t\left[d\left(M_t V_t\right) + M_t D_t dt\right] = 0. \tag{5.30}$$

We can still use equation (5.29) to analyze a security that pays dividends, provided that we interpret V_t as an index whose instantaneous rate of change equals the total return on the security.

Given a process for the stochastic discount factor, we can price any security in the market. Harrison and Kreps (1979) show that we can also work the other way around, and find the stochastic discount factor that is consistent with a set of observed equilibrium prices in the economy. For example, suppose that the only securities in the market are security P_t, whose total return follows the process (5.1), and an instantaneously riskless asset given in (5.2). Further, assume that markets are complete, i.e., that the vector of traded security prices perfectly spans the vector of state variables.

In our example, the complete-markets assumption means that innovations to the state variable and innovations to the risky asset return must be perfectly correlated so that there is only one source of uncertainty in the model (i.e., $dZ_{P,t} = dZ_{S,t}$). In this case, the stochastic discount factor follows a diffusion process with only one diffusion term:

$$\frac{dM_t}{M_t} = \mu_M\left(S,t\right)dt + \sigma_M\left(S,t\right)dZ_{P,t}. \tag{5.31}$$

We can use the martingale property (5.29) to solve for μ_M and σ_M as functions of the drift and diffusion coefficients of security prices. Equation (5.29) implies that $\mathrm{E}_t[d(M_tP_t)] = 0$ and $\mathrm{E}_t[d(M_tB_t)] = 0$. That is, the drift terms of $d(M_tP_t)$ and $d(M_tB_t)$ must be equal to zero. From Itô's Lemma we have

$$\begin{aligned} d(M_tP_t) &= M_tP_t(dP_t + dM_t + \sigma_P\sigma_M dt) \quad (5.32) \\ &= M_tP_t\left[(\mu_P + \mu_M + \sigma_P\sigma_M)dt + (\sigma_P + \sigma_M)dZ_{P,t}\right]. \end{aligned}$$

Thus the martingale property $\mathrm{E}_t[d(M_tP_t)] = 0$ holds if and only if

$$\mu_P + \mu_M + \sigma_P\sigma_M = 0. \quad (5.33)$$

Similarly, for the instantaneously riskless bond we have

$$\begin{aligned} d(M_tB_t) &= M_tB_t(dB_t + dM_t) \quad (5.34) \\ &= M_tB_t\left[(r + \mu_M)dt + \sigma_M dZ_{P,t}\right], \end{aligned}$$

and the martingale property requires

$$r + \mu_M = 0. \quad (5.35)$$

Equations (5.33) and (5.35) define a system of two linear equations with two unknowns, whose unique solution is

$$\mu_M(S,t) = -r(S,t), \quad (5.36)$$

$$\sigma_M(S,t) = -\frac{\mu_P(S,t) - r(S,t)}{\sigma_P(S,t)}. \quad (5.37)$$

That is, the instantaneous expected return on the stochastic discount factor is the negative of the instantaneous interest rate, and the diffusion term is the negative of the price of risk (or Sharpe ratio of the risky asset). Note that if markets are not complete, the innovations to the discount factor will depend on both $dZ_{P,t}$ and $dZ_{S,t}$, and it will not be possible to uniquely identify the drift and diffusion terms of the process for the stochastic discount factor.

It is straightforward to extend this approach to any number of securities. In this case, dP_t/P_t becomes a vector, and $\sigma_{P,t}$ a vector such that $\sigma_{P,t}\sigma'_{P,t} = \Sigma_{P,t}$, where $\Sigma_{P,t}$ is the instantaneous variance-covariance matrix of returns.

5.2.2 Using the stochastic discount factor to solve portfolio and consumption problems

Transforming the dynamic problem into a static problem

The Cox-Huang solution approach uses the properties of the stochastic discount factor under complete markets to solve portfolio and consumption problems. We have defined the optimal portfolio and consumption problem for a long-term investor in Section 5.1.1 as

$$\max_{C,\alpha} \mathrm{E}_0 \left[\int_0^\infty U(C,t)\, dt \right], \tag{5.38}$$

subject to

$$dW = \left[\left(\alpha \left(\mu_P - r \right) + r \right) W - C \right] dt + \alpha W \sigma_P dZ_P, \tag{5.39}$$

and positive initial wealth $W_0 > 0$. Note that by simply reordering terms we can rewrite the intertemporal budget constraint as:

$$\frac{dW + Cdt}{W} = \left[\alpha \mu_P + (1-\alpha) r \right] dt + \alpha \sigma_P dZ_P. \tag{5.40}$$

This expression interprets the dynamic budget constraint (5.39) as the total return on an asset whose price is W_t and that has an instantaneous dividend each period equal to optimal consumption C_t. Under this interpretation, optimally invested wealth W_t must verify

$$W_t = \mathrm{E}_t \left[\int_t^\infty C_s \frac{M_s}{M_t} ds \right]. \tag{5.41}$$

That is, optimally invested wealth is the expected present value of optimal future consumption discounted using the stochastic discount factor. Optimally invested wealth at any time must be able to finance expected consumption under the optimal consumption plan determined at t.

With this reinterpretation of the budget constraint, we can transform the dynamic optimization problem (5.38)-(5.39) into the following problem:

$$\max_{C} \mathrm{E}_0 \left[\int_0^\infty U(C,t)\, dt \right] \tag{5.42}$$

subject to

$$W_0 = \mathrm{E}_0 \left[\int_0^\infty C_t \frac{M_t}{M_0} dt \right], \tag{5.43}$$

where we omit α from the argument of the max operator because we assume that W_0 and C_t in (5.43) denote optimally invested wealth and optimal consumption respectively. Cox and Huang (1989) show that the solution to this problem is equivalent to the solution to problem (5.38)-(5.39). At the same time, (5.42)-(5.43) is a static problem that we can solve using the standard Lagrangian method.

The first-order conditions for the static problem (5.42)-(5.43) are

$$U_C(C) = \ell M_t \Longrightarrow C = U_C^{-1}(\ell M_t), \tag{5.44}$$

where ℓ denotes the Lagrange multiplier, and (5.43)—that is, the budget constraint must hold along the optimal path. Note that ℓ does not have a time subscript; it is a constant determined at time 0. Substituting (5.44) into (5.43) we have

$$W_0 = \mathrm{E}_0\left[\int_0^\infty U_C^{-1}(\ell M_t)\frac{M_t}{M_0}dt\right]. \tag{5.45}$$

We can simplify this expression by defining a new variable

$$X_t = (\ell M_t)^{-1}. \tag{5.46}$$

This definition implies that $M_t/M_0 = X_0/X_t$. It also implies the following dynamics for X_t:

$$\begin{aligned}\frac{dX_t}{X_t} &= -\frac{dM_t}{M_t} + \left(\frac{dM_t}{M_t}\right)^2 \\ &= \left(-\mu_M(S,t) + \sigma_M^2(S,t)\right)dt - \sigma_M(S,t)\,dZ_{P,t},\end{aligned} \tag{5.47}$$

where the first line follows from Itô's Lemma, and the second line follows from (5.31), (5.36) and (5.37). Cox and Huang (1989) give an interesting interpretation of the variable X_t, noting that it is the value of the "growth-optimal portfolio." We showed in Section 2.1.3 that this is the portfolio that maximizes the log return on wealth, and it is the optimal portfolio for an investor with log utility over terminal wealth.[2]

Substituting back into (5.45) and noting that (5.45) must hold at all times, we obtain the following equality for optimally invested wealth:

$$\begin{aligned}W_t &= \mathrm{E}_t\left[\int_t^\infty U_C^{-1}\left(X_s^{-1}\right)\frac{X_t}{X_s}ds\right] \\ &= X_t\,\mathrm{E}_t\left[\int_t^\infty U_C^{-1}\left(X_s^{-1}\right)X_s^{-1}ds\right].\end{aligned} \tag{5.48}$$

[2]This investor chooses a portfolio policy given by $\alpha = (\mu_P - r)/\sigma_P^2 = \sigma_M/\sigma_P$. Substitution of this rule into (5.5)—with $C_t = 0$—gives (5.47).

Given the Markovian structure of the dynamics for X_t and S_t (see [5.3] and [5.47]), this expectation will be some function F of the current value of X. If the process for X_t depends on the state variable, it will be also a function of the current value of S:

$$W_t = F(X, S, t). \tag{5.49}$$

This observation has important implications for the role of time-varying expected returns, variances and covariances on portfolio choice. Note that the process for X depends only on the instantaneous interest rate, $-\mu_M = r$, and on the price of risk, $-\sigma_M = (\mu_P - r)/\sigma_P$, but it does not depend on the expected return on the risky asset μ_P or its instantaneous standard deviation σ_P in isolation. Thus, if both the instantaneous interest rate and the price of risk are constant, optimally invested wealth will not depend on S, even if μ_P and σ_P are functions of S individually. Optimal portfolio choice and consumption will be also independent of the process for the state variable, because they depend on the state variable only indirectly through optimally invested wealth.

Nielsen and Vassalou (2000) show that this result holds generally, regardless of the dimensions of the state vector and the vector of risky assets, and regardless of whether markets are complete or not. They note that the interest rate is the intercept of the instantaneous capital market line, and the price of risk (or Sharpe ratio) is the slope. Their result implies that the only time variation that matters for consumption and portfolio choice is time-variation in the slope and intercept of the instantaneous capital market line.

Solving for optimally invested wealth

We can use the martingale property (5.29) to solve for the function $F(X, S, t)$. The martingale property implies

$$\mathrm{E}_t\left[d(M_t W_t) + M_t C_t dt\right] = \mathrm{E}_t\left[d(M_t F_t) + M_t U_C^{-1}\left(X_t^{-1}\right) dt\right] = 0, \tag{5.50}$$

where the second term on the right-hand side comes from the fact that optimally invested wealth pays an instantaneous "dividend" equal to $C_t dt$. We now show that the expectation (5.50) implies a second-order partial differential equation (PDE) for optimally invested wealth.

To compute the expectation (5.50), we need first to compute $d(M_t F_t)$. By Itô's Lemma, we have that

$$d(M_t F_t) = F_t dM_t + M_t dF_t + dM_t dF_t. \tag{5.51}$$

We have already derived the dynamics for dM_t in (5.51). We can obtain the dynamics for dF by using Itô's Lemma once again:

$$dF = F_X dX + F_S dS + \partial F/\partial t + \frac{1}{2} F_{XX} (dX)^2 + \frac{1}{2} F_{SS} (dS)^2 + F_{XS} dX dS, \tag{5.52}$$

where subindices denote partial derivatives—for example, $F_S = \partial F/\partial S$—with the exception that we write $\partial F/\partial t$ rather than F_t to avoid any confusion with the value of the function F at time t.

Direct substitution of equations (5.51) and (5.52) in (5.50) implies that the argument of the expectation follows an Itô process. Thus the expectation in (5.50) is zero only if the drift of this process is zero. Setting the drift to zero, we obtain a second-order partial differential equation (PDE) for optimally invested wealth:

$$\begin{aligned} & U_C^{-1}\left(X^{-1}\right) + \left(r + \sigma_M^2\right) X F_X + \mu_S F_S + (1-r)\,\partial F/\partial t \\ & + \frac{1}{2}\sigma_M^2 X^2 F_{XX} + \frac{1}{2}\sigma_{S,t}^2 F_{SS} - \sigma_M \sigma_S X F_{XS} + \\ = & \left(\sigma_M^2 F_X X - \sigma_M \sigma_S F_S\right), \end{aligned} \tag{5.53}$$

with boundary condition

$$\lim_{t\to\infty} \mathrm{E}_0 \left[F\left(X, S, t\right)\right] = 0. \tag{5.54}$$

Solving for optimal consumption and portfolio choice

Once we have solved for optimally invested wealth $W_t \equiv F(X, S, t)$, we can easily solve for consumption and portfolio choice. To solve for consumption, we use the first order condition (5.44):

$$\begin{aligned} C_t &= U_C^{-1}\left(\frac{1}{X_t}\right) \\ &= U_C^{-1}\left(\frac{1}{F^{-1}\left(W, S, t\right)}\right), \end{aligned} \tag{5.55}$$

where we have assumed that $F(X, S, t)$ is invertible, so $W_t \equiv F(X, S, t) \Longrightarrow X_t = F^{-1}(W, S, t)$.

To solve for optimal portfolio choice, we simply equate the diffusion terms of the intertemporal budget constraint (5.5) and the equation describing the dynamics of optimally invested wealth (5.52), since both must be the same along the optimal path:

$$\begin{aligned} \alpha F \sigma_P dZ_{P,t} &= -F_X X \sigma_M dZ_{P,t} + F_S \sigma_S dZ_{S,t} \\ &= \left(\frac{\mu_P - r}{\sigma_P} \right) F_X X dZ_{P,t} + F_S \sigma_S dZ_{S,t}. \end{aligned} \tag{5.56}$$

This equation highlights the importance of a complete-markets assumption in the Cox-Huang approach. The left-hand side of this equation depends only on $dZ_{P,t}$, but the right-hand side depends on both $dZ_{P,t}$ and $dZ_{S,t}$. To identify α we need either that $F_S = 0$, or the complete-markets assumption that $dZ_{P,t}$ and $dZ_{S,t}$ are perfectly correlated so that $dZ_{S,t} = dZ_{P,t}$. Since F_S need not be zero in general, solving for α requires that we assume that markets are complete. Under this assumption, the optimal portfolio rule is

$$\alpha = \frac{F_X X}{F} \left(\frac{\mu_P (S,t) - r(S,t)}{\sigma_P^2 (S,t)} \right) + \frac{F_S}{F} \left(\frac{\sigma_S (S,t)}{\sigma_P (S,t)} \right). \tag{5.57}$$

By analogy with equation (5.11) with $\rho_{S,P}(S,t) = 1$, we can easily identify the first component of equation (5.57) with the myopic component of the optimal portfolio rule, and the second component with the hedging component. As Cox and Huang (1989) note, equations (5.11) and (5.57) also allow us to relate the dynamic programming approach and the Cox-Huang approach. Direct comparison of these equations shows that the value function and optimally invested wealth verify the following identities:

$$F_X X = -\frac{J_W}{J_{WW}}, \qquad F_S = -\frac{J_{WS}}{J_{WW}}. \tag{5.58}$$

5.2.3 Our example revisited

We can easily apply these results to the example of optimal portfolio choice with time-varying interest rates and power utility given in section 5.1.2. First, note that the dynamics for the instantaneous interest rate and the return on a long-term bond imply the following process for the stochastic discount factor:

$$\frac{dM_t}{M_t} = -r_t dt + \frac{\lambda}{\sigma_r} dZ_{r,t}. \tag{5.59}$$

This process implies that $dX/X = r + \lambda^2/\sigma_r^2 - \lambda/\sigma_r dZ_r$ and $dXdS \equiv dXdr = -\lambda dt$. With power utility defined over consumption, $U(C_t) = \mathrm{e}^{-\beta t} C_t^{1-\gamma}/(1-\gamma)$. Thus

$$C_t = \mathrm{e}^{-\beta t/\gamma} X_t^{1/\gamma} \equiv U_C^{-1}(X_t^{-1}). \tag{5.60}$$

Substituting these results into (5.53) we obtain:

$$\begin{aligned} 0 &= \mathrm{e}^{-\beta t/\gamma} X^{1/\gamma} - rF + rXF_X + (\kappa_r(\theta_r - r) + \lambda) F_r + \partial F/\partial t \\ &\quad + \frac{1}{2}\frac{\lambda^2}{\sigma_r^2} X^2 F_{XX} + \frac{1}{2}\sigma_r^2 F_{rr} - \lambda X F_{Xr}, \end{aligned} \tag{5.61}$$

subject to the boundary condition $\lim_{t\to\infty} \mathrm{E}_0 [F(X, r, t)] = 0$.

Wachter (1999) shows that this equation has a solution of the form

$$F(X, r, t) = \mathrm{e}^{-\beta t/\gamma} X^{\frac{1}{\gamma}} \lim_{\tau\to\infty} \int_0^{\tau - t} \Psi(r_t, s)\, ds, \tag{5.62}$$

where

$$\Psi(r_t, s) = \exp\left\{ \frac{(1-\gamma) b(s)}{\gamma} r_t + A(s) - \frac{\beta}{\gamma} s \right\}, \tag{5.63}$$

and $A(s)$ is a function of s. Wachter (1999) notes that the limit of the function $\Psi(r_t, s)$ as $\gamma \to \infty$ is the price of a zero-coupon bond with maturity s that pays one unit of consumption at maturity: $\lim_{\gamma\to\infty} \Psi(r_t, s) = P(r, s) = \exp\{-b(s) r_t + A(s)\}$. This has important implications for the interpretation of the optimal consumption and portfolio rules.

The optimal rules can be found by direct substitution of $F(X, r, t)$ into (5.55) and (5.57). The optimal consumption rule is

$$\frac{C}{W} = \frac{U_C^{-1}(X^{-1})}{F} = \left(\lim_{\tau\to\infty} \int_0^{\tau - t} \Psi(r_t, s)\, ds \right)^{-1}. \tag{5.64}$$

Since $\lim_{\gamma\to\infty} \Psi(r_t, s)$ is the price of a zero coupon bond with maturity s, the limit as $\gamma \to \infty$, of the integral expression in (5.64) must be the price of a coupon bond with maturity τ. Therefore, as $\tau \to \infty$ and $\gamma \to \infty$, this expression converges to the price of a real consol bond that pays one unit of consumption each period. Optimally invested wealth for an infinitely risk averse investor is equal to the value of a real consol bond that pays C units of consumption each period.

For an investor who is not infinitely risk averse, a similar interpretation is still possible. Wachter (1999) shows that $\Psi(r_t, s)$ is the current value of one unit of consumption s periods ahead for an investor with relative risk aversion coefficient γ who is maximizing utility over consumption s periods ahead. Therefore, the integral expression in (5.64) is the value for this investor of one unit of consumption per period in the future.

The optimal portfolio allocation to long-term bonds is

$$\begin{aligned}\alpha &= \frac{XF_X}{F}\frac{\lambda}{\sigma_r^2 b\left(T-t\right)} - \frac{F_r}{F}\frac{1}{b\left(T-t\right)} \qquad (5.65)\\ &= \frac{1}{\gamma}\frac{\lambda}{b\left(T-t\right)\sigma_r^2} + \left(1-\frac{1}{\gamma}\right)\frac{1}{b\left(T-t\right)}\lim_{\tau\to\infty}\frac{\int_0^{\tau-t}\Psi\left(r_t,s\right)b\left(s\right)ds}{\int_0^{\tau-t}\Psi\left(r_t,s\right)ds},\end{aligned}$$

where the first term of α is the myopic demand for long-term bonds and the second term is the intertemporal hedging component. Given our earlier interpretation of $\int_0^{\tau-t}\Psi(r_t,s)$, it is easy to see that the ratio of integrals in the intertemporal hedging component measures the modified duration of optimal consumption. Wachter (1999) shows that as $\gamma\to\infty$, α converges to a position in the available zero-coupon bond that, in combination with the instantaneously riskless asset, replicates the payments on a real consol bond.

Equation (5.62) writes the solution of the PDE (5.61) in integral form. To evaluate this function, it would be necessary to do numerical integration. However, it is possible to find an approximate analytical solution to this PDE that does not require numerical integration. This solution is identical to the solution we obtained using dynamic programming. To see this, note that the first term of equation (5.61) is simply the consumption-wealth ratio. Thus we can approximate this ratio using the same log-linear approximation as in section 5.1.3. First, we substitute $h_0+h_1(c_t-f_t)$ for the first term in the equation. Next, we guess that

$$F\left(X,r,t\right)=\mathrm{e}^{-\beta t/\gamma}X^{\frac{1}{\gamma}}\exp\left\{C_0+C_1r_t\right\}, \qquad (5.66)$$

and we use this guess to compute all the expressions in the PDE (5.61) involving F or its derivatives. We also note that this guess implies $c_t-f_t=-C_0-C_1r_t$. It is straightforward to see that substitution of the guess into the approximated PDE leads to the same approximate analytical solution for the optimal portfolio policy and the optimal consumption-wealth ratio that we obtained before in equations [5.25) and [5.26].

5.3 Recursive Utility in Continuous Time

In Chapter 2 we introduced recursive Epstein-Zin preferences as a way to generalize the standard, time-separable power utility model to separate relative risk aversion from the elasticity of intertemporal substitution of consumption. Duffie and Epstein (1992a, 1992b) and Fisher and Gilles (1998) derive a continuous-time analogue of the Epstein-Zin utility function. We adopt the Duffie and Epstein (1992b) parameterization of recursive utility:

$$J_t = \mathrm{E}_t \left[\int_t^\infty \mathrm{f}\left(C_s, J_s\right) ds \right], \tag{5.67}$$

where $\mathrm{f}(C, J)$ is a normalized aggregator of current consumption and continuation utility that takes the form

$$\mathrm{f}\left(C, J\right) = \frac{\beta}{1 - \frac{1}{\psi}} \left(1 - \gamma\right) J \left[\left(\frac{C}{\left(\left(1 - \gamma\right) J\right)^{\frac{1}{1-\gamma}}} \right)^{1 - \frac{1}{\psi}} - 1 \right]. \tag{5.68}$$

Here $\beta > 0$ is the rate of time preference, $\gamma > 0$ is the coefficient of relative risk aversion, and $\psi > 0$ is the elasticity of intertemporal substitution. Power utility obtains from (5.68) by setting $\psi = 1/\gamma$.

The normalized aggregator $\mathrm{f}(C, J)$ takes the following form when $\psi \to 1$:

$$\mathrm{f}\left(C, J\right) = \beta \left(1 - \gamma\right) J \left[\log\left(C\right) - \frac{1}{1 - \gamma} \log\left(\left(1 - \gamma\right) J\right) \right]. \tag{5.69}$$

Duffie and Epstein (1992a, 1992b) show that the Bellman Principle of Optimality applies to recursive utility. From a computational perspective, the only difference from the standard additive utility case is that we need to substitute the normalized aggregator $\mathrm{f}(C, J)$ for the instantaneous utility function $U(C)$ in the Bellman equation (5.6).

5.3.1 Our example revisited once more: An exact solution with unit elasticity of intertemporal substitution.

This section derives a solution to the model with stochastic interest rates under recursive utility. This is the continuous-time counterpart of the model with inflation-indexed bonds that we solved in section 3.2. We argued in chapters 3 and 4 that our discrete-time approximate analytical solution is exact in continuous time with diffusions for asset prices, provided that the elasticity of intertemporal substitution is one. We now prove this claim.

The Bellman equation for the recursive-utility model with $\psi = 1$ is identical to (5.16), except that we substitute (5.69) for the instantaneous power utility of consumption $\mathrm{e}^{-\beta t} C^{1-\gamma}/(1-\gamma)$. The first-order condition for consumption (the envelope condition) becomes

$$C = \beta \left(1 - \gamma\right) \frac{J}{J_W}, \tag{5.70}$$

and the first-order condition for α is identical to (5.18). We now guess that the solution to the Bellman equation has the form

$$J(W,r,t) = I(r,t)\frac{W^{1-\gamma}}{1-\gamma}. \tag{5.71}$$

Substitution of this guess and the first-order conditions into the Bellman equation lead to the following ordinary differential equation (ODE):

$$\begin{aligned} 0 &= -\frac{\beta}{1-\gamma}\log I + \left(\beta\log\beta - \beta + \frac{\lambda^2}{2\gamma\sigma_r^2} + r\right) \\ &\quad + \frac{\sigma_r^2}{2\gamma}\left(\frac{I_r}{I}\right)^2 + \left(\frac{\kappa_r}{1-\gamma}(\theta_r - r) - \frac{\lambda}{\gamma}\right)\frac{I_r}{I} + \frac{\sigma_r^2}{2(1-\gamma)}\frac{I_{rr}}{I}. \end{aligned} \tag{5.72}$$

This equation has an exact solution of the form

$$I(r,t) = \exp\{C_0 + C_1 r_t\}, \tag{5.73}$$

where $C_1 = (1-\gamma)/(\beta+\kappa_r)$, and C_0 is given in the Appendix. This solution implies a constant consumption-wealth ratio equal to β, and an optimal portfolio rule given by (5.26) with $h_1 = \beta$.

When the elasticity of intertemporal substitution is not equal to one, we can still obtain an approximate analytical solution along the lines of the solution we presented in section 5.1. The Bellman equation is once again identical to (5.16), except that we substitute (5.68) for the instantaneous power utility of consumption. The first-order condition for consumption is now

$$C = \frac{\left[(1-\gamma)J_t\right]^{\frac{1-\gamma\psi}{1-\gamma}}\beta^{\psi}}{J_W^{-\psi}}, \tag{5.74}$$

and the first-order condition for portfolio choice is again identical to (5.18). We guess a solution of the form

$$J(W,r,t) = H(r,t)^{-\frac{1-\gamma}{1-\psi}}\frac{W^{1-\gamma}}{1-\gamma}, \tag{5.75}$$

which leads to the following non-homogeneous ODE:

$$\begin{aligned} 0 &= \frac{\gamma}{1-\gamma}\beta^{\psi}H^{-1} + \left(-\frac{(1-\psi)\lambda^2}{2(1-\gamma)\sigma_r^2} - \frac{\gamma\psi\beta}{1-\gamma} - \frac{(1-\psi)\gamma}{1-\gamma}r\right) \\ &\quad + \left(\frac{\gamma\kappa_r}{1-\gamma}(\theta_r - r) - \lambda\right)\left(\frac{H_r}{H}\right) - \left(\frac{1}{1-\psi} + \frac{\gamma}{1-\gamma}\right)\frac{\sigma_r^2}{2}\left(\frac{H_r}{H}\right)^2 \\ &\quad + \frac{\gamma\sigma_r^2}{2(1-\gamma)}\left(\frac{H_{rr}}{H}\right). \end{aligned} \tag{5.76}$$

Equation (5.76) reduces to equation (5.20) when $\psi = 1/\gamma$, i.e., when recursive utility reduces to time-additive power utility. We can find an approximate analytical solution to this equation using the same approach as in section 5.1. Once again, the envelope condition (5.74) implies that first term of the equation, $\beta^{\psi} H^{-1}$, is the optimal consumption-wealth ratio. Using the loglinear approximation $\beta^{\psi} H^{-1} \approx h_0 + h_1(c_t - w_t)$, the resulting ODE has a solution of the form $H_t = \exp\{C_0 + C_1 r_t\}$, with $C_1 = -(1 - \psi)/(h_1 + \kappa_1)$. The optimal log consumption-wealth ratio is given by

$$c_t - w_t = \psi \log \beta - C_0 + \frac{1-\psi}{h_1 + \kappa_r} r_t, \tag{5.77}$$

and the optimal portfolio rule is identical to the optimal rule under power utility (see equation [5.26]). Thus, the optimal portfolio rule depends on the investor's willingness to substitute consumption intertemporally only indirectly, through the parameter h_1 that determines the mean log consumption wealth-ratio.

5.4 Should Long-Term Investors Hedge Stock Return Volatility Risk?

The continuous-time approach that we have just presented is especially helpful in showing how long-term investors should react to time-varying risk. Motivated by empirical evidence, Chapters 3 and 4 examined the relevance of time-variation in interest rates and expected excess bond and stock returns for long-term portfolio choice. There is equally strong empirical evidence that the volatility of stock returns is time varying. Partial surveys of the enormous literature on time-varying volatility are given by Bollerslev, Chou, and Kroner (1992), Hentschel (1995), Ghysels, Harvey, and Renault (1996), and Campbell, Lo, and MacKinlay (1997, Chapter 12).

Chacko and Viceira (1999) explore the implications of changing volatility for long-term portfolio choice. They assume that the only source of time-variation in investment opportunities is time-variation in instantaneous precision, the inverse of the instantaneous variance of stock returns. Their model writes y_t for instantaneous precision, and equations (5.1)-(5.3) become

$$\frac{dP_t}{P_t} = \mu_P dt + \sqrt{\frac{1}{y_t}} dZ_{P,t}, \tag{5.78}$$

$$\frac{dB_t}{B_t} = r dt \tag{5.79}$$

$$dy_t = \kappa_y (\theta_y - y_t) dt + \sigma_y dZ_{y,t}, \tag{5.80}$$

and $dZ_{P,t} dZ_{y,t} = \rho_{Py} dt$. Thus precision follows a mean-reverting process correlated with stock returns, with long-term mean equal to θ_y and reversion parameter $\kappa_y > 0$.[3] This modeling choice implies that the ratio of the mean expected excess stock return to the variance of stock returns, which determines the myopic portfolio, is linear in the state variable (precision). However, the Sharpe ratio is not a linear function of the state variable, but a square-root function. Thus this model is not mathematically equivalent to the model we discussed earlier in this chapter with a linear, mean-reverting process for the expected excess return; that model implies both a linear Sharpe ratio and a linear ratio of mean excess return to variance.

Chacko and Viceira (1999) note that the parameterization of the precision process implies a mean-reverting process for instantaneous volatility $v_t = 1/y_t$. The process for v_t can be found by applying Ito's Lemma to (5.80):

$$\frac{dv_t}{v_t} = \kappa_v (\theta_v - v_t) dt - \sigma_y \sqrt{v_t} dZ_{y,t}, \tag{5.81}$$

where $\theta_v = (\theta_y - \sigma_y^2/\kappa_y)^{-1}$ and $\kappa_v = \kappa_y/\theta_v$. Equation (5.81) implies that proportional changes in volatility are correlated with stock returns, with instantaneous correlation

$$\text{Corr}_t(\frac{dv_t}{v_t}, \frac{dS_t}{S_t}) = -\rho_{Py}. \tag{5.82}$$

Equation (5.81) can capture the main stylized empirical facts about stock return volatility: Stock return volatility appears to be mean-reverting and negatively correlated with stock returns. Moreover, proportional changes in volatility are more pronounced in times of high volatility than in times of low volatility. Table 5.1, taken from Chacko and Viceira (2000), shows estimates of equations (5.78)-(5.81) for US monthly stock returns from January 1926 through December 1997, and annual stock returns from 1871 through 1997.[4] Standard errors appear in parenthesis, and parameter estimates are annualized to facilitate their interpretation.

[3] In order to satisfy standard integrability conditions, we assume that $2\kappa_y\theta_y > \sigma_y^2$.

[4] These estimates are obtained using the Spectral Generalized Method of Moments of

Table 5.1: Stochastic Volatility Model Estimation

	1926.01 - 1997.12	**1871 - 1997**
$\boldsymbol{\mu - r}$	.0799	.0841
	(.0238)	(.0370)
$\boldsymbol{\kappa}$	.3413	.0426
	(.3114)	(.0445)
$\boldsymbol{\theta}$	27.7088	24.7718
	(1.8153)	(12.6946)
$\boldsymbol{\sigma}$	.6512	1.1786
	(.4855)	(.7065)
$\boldsymbol{\rho}$	.5355	.3708
	(.2381)	(.3769)

The estimates in Table 5.1 imply a mean excess stock return of about 8% per year, and an unconditional standard deviation of stock returns around 20% per year. The instantaneous correlation between shocks to volatility and stock returns ($-\rho_{Py}$) is negative and relatively large—almost -54% in the monthly sample and -37% in the annual sample. The estimate of the mean-reversion parameter κ_y implies a half-life of a precision shock of about 2 years in the monthly sample,and about 16 years in the annual sample. French, Schwert and Stambaugh (1987) and Campbell and Hentschel (1990) have also found a relatively slow decay rate for volatility shocks in low-frequency data. This slow reversion to the mean in low-frequency data contrasts with the fast decay rate detected in high-frequency data by Andersen, Benzoni and Lund (1998) and Chacko and Viceira (1999).[5]

We assume that investors' preferences are described by the Duffie-Epstein recursive utility function (5.68)-(5.69), and the intertemporal budget constraint is given by

$$dW_t = [\alpha_t(\mu - r)W_t + rW_t - C_t]dt + \alpha_t W_t \sqrt{\frac{1}{y_t}} dZ_{P,t}. \tag{5.83}$$

Chacko and Viceira (1999) present an exact solution for the case with unit elasticity of intertemporal substitution, and an approximate solution for all other cases. They show that, consistent with the results of chapters 3 and 4, empirically the effect of intertemporal substitution on the optimal portfolio rule is negligible. Thus we present only their exact solution with $\psi = 1$. In this case, (5.69) and (5.83) imply the following Bellman equation:

$$\begin{aligned} 0 \quad = \quad & \sup_{\pi, C} \Big\{ \mathrm{f}(C, J) + [\alpha(\mu - r)W + rW - C]J_W + \frac{1}{2}\alpha^2 W^2 J_{WW} \frac{1}{y} \\ & + \kappa_y(\theta_y - y)J_y + \frac{1}{2}\sigma_y^2 J_{yy} y + \rho_{Py}\sigma_y \alpha W J_{Xy} \Big\}, \end{aligned} \tag{5.84}$$

where $\mathrm{f}(C, J)$ is given in (5.69) and subscripts on J denote partial derivatives.

Chacko and Viceira (1999) and Singleton (1997). This estimation method is essentially a generalized method of moments based on the characteristic function of the stock return process. The source of the monthly data is CRSP, while the source of the annual data is Shiller (1989) and subsequent updates.

[5]Chacko and Viceira (1999) estimate the half-life of a shock to precision to be about 3 months in weekly data for the period 1962-1998. These results suggest the presence of high frequency and low frequency (or long-memory) components in stock market volatility. Chacko and Viceira (1998) show that a model of multiple additive components in stock return volatility, each one operating at a different frequency, generates a similar pattern in the estimates of κ_y when stock returns are sampled at different frequencies.

The first-order condition for consumption is identical to (5.84), and the first-order condition for portfolio choice is

$$\pi_t = \frac{1}{-W\,J_{WW}/J_W}\left(\mu - r\right)y - \frac{J_{Wy}}{W\,J_{WW}}\rho_{Py}\sigma_y y. \tag{5.85}$$

Substitution of the first-order conditions into the Bellman equation, and a guess for the value function of the form $J(W,y,t) = I(y,t)W_t^{1-\gamma}/(1-\gamma)$ yield the following ODE:

$$\begin{aligned} 0 \;=\; & \left(\beta\log\beta - \beta - \frac{\beta\gamma}{1-\gamma}\log\left(1-\gamma\right) + \frac{\left(\mu-r\right)^2}{2\gamma}y + r\right) \\ & -\frac{1}{1-\gamma}\beta\log I + \left(\frac{\rho_{Py}\sigma_y\left(\mu-r\right)}{\gamma}y + \frac{\kappa_y}{1-\gamma}(\theta_y - y)\right)\left(\frac{I_y}{I}\right) \\ & +\frac{\rho^2\sigma^2}{2\gamma}y\left(\frac{I_y}{I}\right)^2 + \frac{\sigma^2}{2\left(1-\gamma\right)}y\left(\frac{I_{yy}}{I}\right). \end{aligned} \tag{5.86}$$

Equation (5.86) has an exact solution of the form $I = \exp\{C_0 + C_1 y_t\}$ that leads to two algebraic equations for C_0 and C_1 given in the Appendix. This solution implies the following optimal portfolio rule:

$$\alpha_t = \frac{1}{\gamma}\left(\mu - r\right)y_t + \left(1 - \frac{1}{\gamma}\right)\rho_{Py}\sigma_y\widetilde{C}_1 y_t, \tag{5.87}$$

where $\widetilde{C}_1 = C_1/(1-\gamma) > 0$.

The optimal portfolio demand for stocks has two components. The first one is the myopic demand, that depends only on the risk premium multiplied by the inverse of the relative risk aversion coefficient and current volatility. The second component is the intertemporal hedging demand. The sign of this demand depends on the sign of the correlation between unexpected returns and changes in volatility $(-\rho_{PY})$ and the sign of $(1-1/\gamma)$. Table 5.1 shows that empirically the correlation ρ_{PY} is negative, which implies that investors with $\gamma > 1$ have a negative intertemporal hedging demand for stocks.

Table 5.2 reports the optimal portfolio allocations to stocks implied by the process estimates shown in Table 5.1. The table assumes a rate of time preference (β) equal to 6% annually; when $\psi = 1$, this is also the optimal, constant consumption-wealth ratio. For each sample period, the table has two columns. The first column ("Mean") reports the mean percentage allocation to stocks, and the second column ("Ratio") reports the percentage ratio of the hedging demand to the myopic demand, which is constant in

Table 5.2: Mean Optimal Percentage Allocation to Stocks and Percentage Hedging Demand Over Myopic Demand

R.R.A.	**1926.01 - 1997.12**		**1871 - 1997**	
	Mean	Ratio	Mean	Ratio
1.00	221.39	0.00	208.33	0.00
1.50	145.93	-1.13	131.87	-5.05
2.00	108.84	-1.67	96.79	-7.08
4.00	53.98	-2.47	47.02	-9.72
10.0	21.49	-2.93	18.52	-11.12
20.0	10.73	-3.09	9.21	-11.57
40.0	5.36	-3.16	4.59	-11.78

(5.87). This ratio tells us the reduction in portfolio demand due to hedging considerations.

Table 5.2 shows that the estimated volatility process implies a small impact of time-variation in volatility on the optimal portfolio demand for stocks. In the monthly sample hedging demand reduces the demand for stocks by at most 3.2%, for highly risk averse investors with $\gamma = 40$, and in the annual sample the reduction in demand is at most 12%. This impact is relatively modest when compared with the effect of time-variation in interest rates or risk premia on the portfolio demand for stocks.

The percentage reduction in the stock demand implied by the annual estimates is at least three times larger than the reduction implied by the monthly estimates. The parameters causing this difference must be the reversion parameter (κ_y) and the correlation between shocks to volatility and stock returns (ρ_{Py}), because these are the only parameters whose magnitude is significantly different across samples. The reversion (or persistence) parameter affects the optimal portfolio demand for stocks through the coefficient $\widetilde{C}_1$. Chacko and Viceira (1999) show that the absolute size of the hedging demand for stocks is increasing in the persistence of volatility shocks when $\gamma > 1$.

Figures 5.1 and 5.2 explore the effects of each parameter on the ratio of hedging demand to myopic demand. Figure 5.1 plots the ratio of hedging

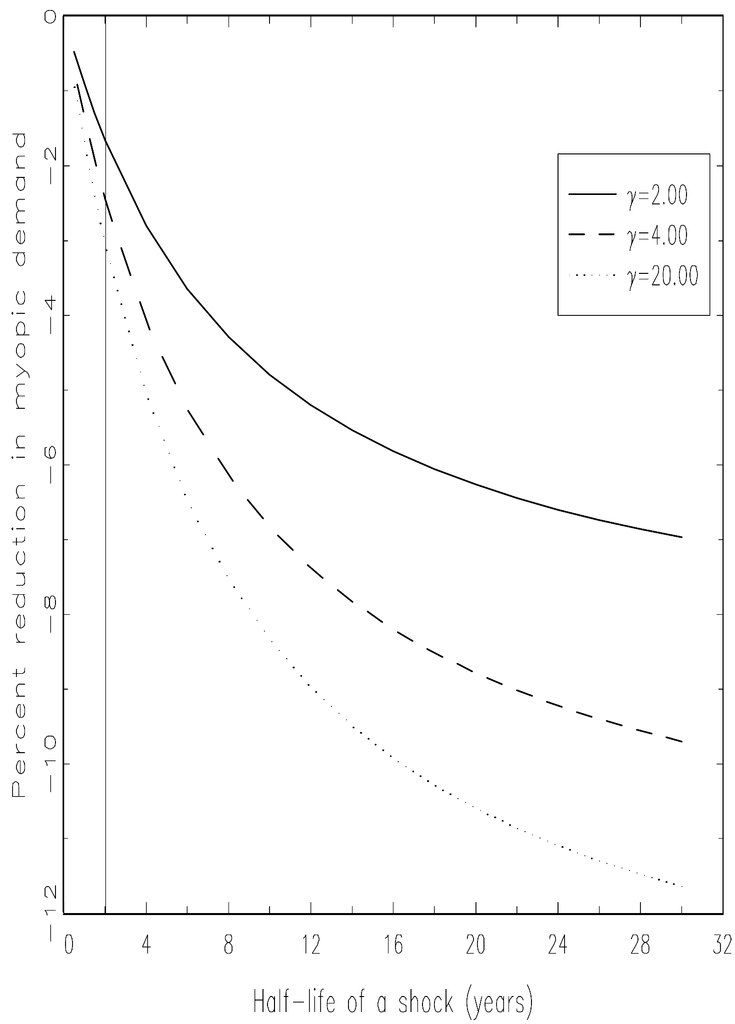

Figure 5.1: Effect of persistence on portfolio demand

demand to myopic demand for values of κ_y implying a half-life of a shock between 6 months and 30 years, holding constant the other parameters at the values implied by the monthly dataset. Figure 5.2 repeats the experiment, this time varying the correlation coefficient and holding constant the other parameters. The vertical line in each plot intersects the horizontal axis at the parameter value implied by the monthly dataset.

Figures 5.1 and 5.2 suggest that hedging demand is more sensitive to the persistence of volatility shocks than to the correlation between volatility shocks and stock returns. Figure 5.1 shows that increasing persistence produces a noticeable reduction on portfolio demand, even for investors with low coefficients of relative risk aversion. For example, an investor with $\gamma = 4$ would reduce her myopic demand by approximately 10% instead of 2.5% if the half-life of a shock were 10 years instead of 2 years. By contrast, the effect of changing the correlation is much smaller. Even if the correlation between unexpected returns and shocks to volatility were -1, hedging demand would not reduce myopic demand by more than 6% for an investor with $\gamma = 20$.

5.5 Conclusion

This chapter has examined the solution to dynamic asset allocation problems in a continuous-time framework. We have linked the approximate solution methodology used in this book to the vast literature on continuous-time portfolio choice.

Continuous-time methods are particularly suitable for modelling time-variation in volatility, so this chapter has explored the implications of volatility movements for asset demand. Empirically, increases in stock market volatility tend to persist for some time, and they are often associated with low realized excess stock returns. Short-term investors should respond by reducing the allocation to equities when volatility increases. Long-term investors should go further. The persistence of volatility shocks, and the negative correlation of these shocks with realized excess stock returns, suggest that long-term investors should hedge volatility risk by reducing their allocation to equities. However, shocks to volatility in the US stock market do not seem to be sufficiently persistent and negatively correlated with stock returns to justify a large negative intertemporal hedging portfolio demand for stocks. When compared to the size of intertemporal hedging demands induced by changes in interest rates and risk premia, the negative intertemporal hedging demand created by time-varying risk is relatively modest.

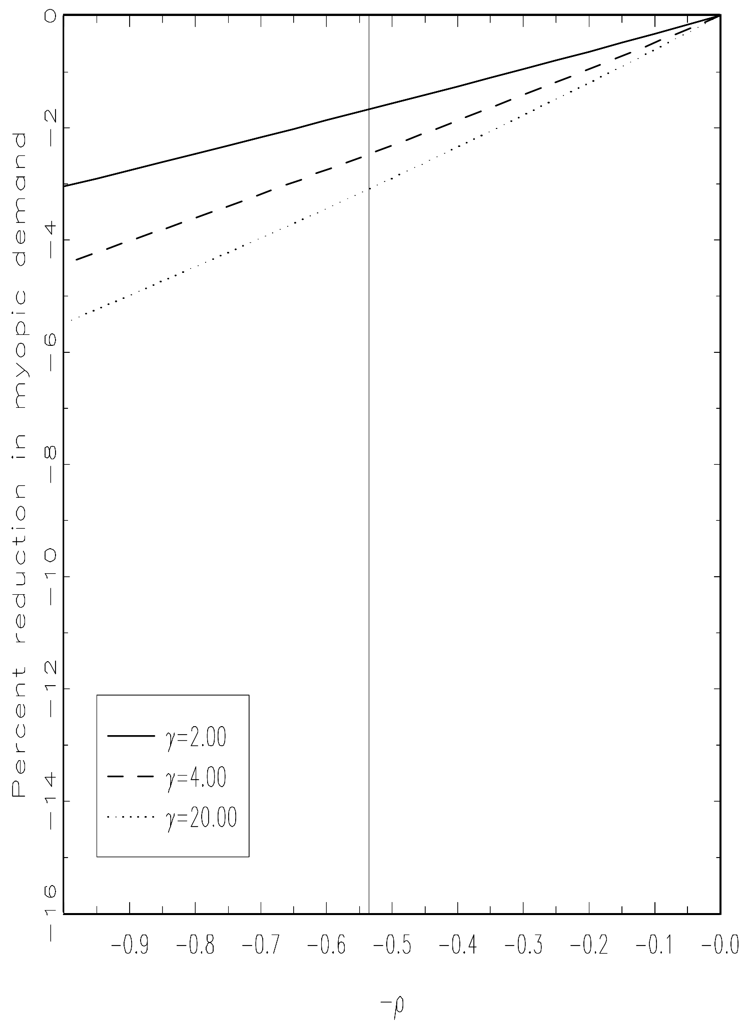

Figure 5.2: Effect of correlation on portfolio demand

A limitation of the empirical analysis in this chapter is that we have assumed constant expected excess stock returns when studying volatility. A fully general model would allow a set of state variables to shift both the equity premium and stock market volatility jointly. As we have noted, hedging demands would then depend on the implied process for the Sharpe ratio. Authors such as Campbell (1987), Harvey (1989, 1991), Glosten, Jagannathan, and Runkle (1993), and Ait-Sahalia and Brandt (2000) have modelled time-varying returns and volatility jointly. These studies typically find that the effects of state variables on expected returns are stronger than their effects on volatility, which suggests that the negative hedging demand associated with volatility risk will be modest even in a framework that combines time-varying volatility with the time-varying returns modelled in Chapter 4.

References

Aït-Sahalia, Yacine and Michael Brandt, 2000, "Variable Selection for Portfolio Choice", unpublished paper, Princeton University and University of Pennsylvania.

Balduzzi, Perluigi and Anthony Lynch, 1999, "Transaction Costs and Predictability: Some Utility Cost Calculations", *Journal of Financial Economics* 52, 47–78.

Barberis, Nicholas C., 2000, "Investing for the Long Run When Returns Are Predictable", *Journal of Finance* 55, 225–264.

Bekaert, Geert, Robert J. Hodrick, and David A. Marshall, 1997, "On Biases in Tests of the Expectations Hypothesis of the Term Structure of Interest Rates", *Journal of Financial Economics* 44, 309–348.

Brandt, Michael, 1999, "Estimating Portfolio and Consumption Choice: A Conditional Euler Equations Approach", *Journal of Finance* 54, 1609–1645.

Breeden, Douglas, 1979, "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities", *Journal of Financial Economics* 7, 265–296.

Brennan, Michael J., 1998, "The Role of Learning in Dynamic Portfolio Decisions", *European Finance Review* 1, 295–306.

Brennan, Michael J., Eduardo S. Schwartz, and Ronald Lagnado, 1996, "The Use of Treasury Bill Futures in Strategic Asset Allocation Programs", Finance Working Paper 7-96, Anderson Graduate School of Management, UCLA, Los Angeles, CA.

Brennan, Michael J., Eduardo S. Schwartz, and Ronald Lagnado, 1997, "Strategic Asset Allocation", *Journal of Economic Dynamics and Control* 21, 1377–1403.

Brennan, Michael J. and Y. Xia, 1998, "Resolution of a Financial Puzzle", unpublished paper, Anderson Graduate School of Management, UCLA, Los Angeles, CA.

Brennan, Michael J. and Y. Xia, 1999, "Assessing Asset Pricing Anomalies", unpublished paper, Anderson Graduate School of Management, UCLA, Los Angeles, CA.

Campbell, John Y., 1987, "Stock Returns and the Term Structure", *Journal of Financial Economics* 18, 373–399.

Campbell, John Y., 1991, "A Variance Decomposition for Stock Returns", Economic Journal 101, 157–179.

Campbell, John Y., 1993, "Intertemporal Asset Pricing without Consumption Data", *American Economic Review* 83, 487–512.

Campbell, John Y., 1996, "Understanding Risk and Return", *Journal of Political Economy* 104, 298–345.

Campbell, John Y., 1999, "Asset Prices, Consumption, and the Business Cycle", Chapter 19 in John Taylor and Michael Woodford eds. *Handbook of Macroeconomics Vol. 1*, North-Holland, Amsterdam.

Campbell, John Y., 2000, "Asset Pricing at the Millennium", *Journal of Finance.*

Campbell, John Y., Yeung Lewis Chan, and Luis M. Viceira, 2000, "A Multivariate Model of Strategic Asset Allocation", unpublished paper, Harvard University.

Campbell, John Y., Joao Cocco, Francisco Gomes, Pascal Maenhout, and Luis Viceira, 1998, "Stock Market Mean Reversion and the Optimal Equity Allocation of a Long-Lived Investor", unpublished paper, Harvard University.

Campbell, John Y., Andrew W. Lo and A. Craig MacKinlay, 1997, *The Econometrics of Financial Markets*, Princeton University Press, Princeton, NJ.

Campbell, John Y. and Hyeng Keun Koo, 1997, "A Comparison of Numerical and Analytical Approximate Solutions to an Intertemporal Consumption Choice Problem", *Journal of Economic Dynamics and Control* 21, 273–295.

Campbell, John Y. and Robert J. Shiller, 1988, "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors", *Review of Financial Studies* 1, 195–228.

Campbell, John Y. and Robert J. Shiller, 1991, "Yield Spreads and Interest Rates: A Bird's Eye View", *Review of Economic Studies* 58, 495–514.

Campbell, John Y. and Robert J. Shiller, 1996, "A Scorecard for Indexed Government Debt", in Ben S. Bernanke and Julio Rotemberg eds. *National Bureau of Economic Research Macroeconomics Annual* 11, 155–197.

Campbell, John Y. and Luis M. Viceira, 1999, "Consumption and Portfolio Decisions when Expected Returns are Time Varying", *Quarterly Journal of Economics* 114, 433–495.

Campbell, John Y. and Luis M. Viceira, 2001, "Who Should Buy Long-Term Bonds?", forthcoming *American Economic Review.*

Canner, Niko, N. Gregory Mankiw, and David N. Weil, 1997, "An Asset Allocation Puzzle", *American Economic Review* 87, 181–191.

Clarida, Richard, Jordi Gali, and Mark Gertler, 1998, "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory", NBER Working Paper No. 6442.

Cocco, Joao, Francisco Gomes, and Pascal Maenhout, 1998, "Consumption and Portfolio Choice over the Life Cycle", unpublished paper, Harvard University.

Cochrane, John H., 2001, *Asset Pricing*, forthcoming Princeton University Press.

Cochrane, John H. and Lars P. Hansen, 1992, "Asset Pricing Explorations for Macroeconomics", in Olivier J. Blanchard and Stanley Fischer eds. *National Bureau of Economic Research Macroeconomics Annual.*

Cox, John C. and Chi-fu Huang, 1989, "Optimal Consumption and Portfolio Policies when Asset Prices Follow a Diffusion Process", *Journal of Economic Theory* 39, 33–83.

Cvitanić, J. and Karatzas, I., 1992, "Convex Duality in Constrained Portfolio Optimization", *Annals of Applied Probability*, 767–818.

Dammon, Robert M., Chester S. Spatt, and Harold H. Zhang, 1999, "Optimal Consumption and Investment with Capital Gains Taxes", unpublished paper, Carnegie Mellon University.

Epstein, Lawrence and Stanley Zin, 1989, "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework", *Econometrica* 57, 937–69.

Epstein, Lawrence and Stanley Zin, 1991, "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Investigation", *Journal of Political Economy* 99, 263–286.

Fama, Eugene and Kenneth French, 1988a, "Dividend Yields and Expected Stock Returns", *Journal of Financial Economics* 22, 3—27.

Fama, Eugene and Kenneth French, 1988b, "Permanent and Temporary Components of Stock Prices", *Journal of Political Economy* 96, 246–273.

Fama, Eugene and Kenneth French, 1989, "Business Conditions and Expected Returns on Stocks and Bonds", *Journal of Financial Economics* 25, 23–49.

Fama, Eugene and G. William Schwert, 1977, "Asset Returns and Inflation", *Journal of Financial Economics* 5, 115–146.

Giovannini, Alberto and Philippe Weil, 1989, "Risk Aversion and Intertemporal Substitution in the Capital Asset Pricing Model", NBER Working Paper 2824, National Bureau of Economic Research, Cambridge, MA.

Glosten, Lawrence R., Ravi Jagannathan, and David Runkle, 1993, "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks", *Journal of Finance* 48, 1779–1801.

Goetzmann, William N. and Philippe Jorion, 1993, "Testing the Predictive Power of Dividend Yields", *Journal of Finance* 48, 663–679.

Gong, Frank F. and Eli Remolona, 1996, "A Three-Factor Econometric Model of the US Term Structure", unpublished paper, Federal Reserve Bank of New York.

Grossman, Sanford J. and Robert J. Shiller, 1981, "The Determinants of the Variability of Stock Market Prices", *American Economic Review* 71, 222–227.

Hansen, Lars P. and Ravi Jagannathan, 1991, "Restrictions on Intertemporal Marginal Rates of Substitution Implied by Asset Returns", *Journal of Political Economy* 99, 225-262.

Hansen, Lars P. and Kenneth J. Singleton, 1983, "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns", *Journal of Political Economy* 91, 249–268.

Harvey, Campbell, 1989, "Time-Varying Conditional Covariances in Tests of Asset Pricing Models", *Journal of Financial Economics* 22, 305–334.

Harvey, Campbell, 1991, "The World Price of Covariance Risk", *Journal of Finance* 46, 111–157.

Hodrick, Robert J., 1992, "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement", *Review of Financial Studies* 5, 357–386.

Kandel, Shmuel, and Robert Stamburgh, 1987, "Long Horizon Returns and Short Horizon Models", CRSP Working Paper No.222, University of Chicago.

Kim, Tong Suk and Edward Omberg, 1996, "Dynamic Nonmyopic Portfolio Behavior", *Review of Financial Studies* 9, 141–161.

Kreps, David and Edward Porteus, 1978, "Temporal Resolution of Uncertainty and Dynamic Choice Theory", *Econometrica* 46, 185–200.

Kritzman, Mark P., 2000, *Puzzles of Finance: Six Practical Problems and Their Remarkable Solutions*, John Wiley, New York, NY.

Liu, Jun, 1998, "Portfolio Choice in Stochastic Environments", unpublished paper, Stanford University.

Lucas, Robert E. Jr., 1978, "Asset Prices in an Exchange Economy", *Econometrica* 46, 1429–1446.

Lynch, Anthony W., 1999, "Portfolio Choice and Equity Characteristics: Characterizing the Hedging Demands Induced by Return Predictability", unpublished paper, New York University.

Markowitz, Harry, 1952, "Portfolio Selection", *Journal of Finance* 7, 77–91.

McCulloch, J. Huston and Heon-Chul Kwon, 1993, "US Term Structure Data 1947–1991", Working Paper 93-6, Ohio State University.

Mehra, Rajnish and Edward C. Prescott, 1985, "The Equity Premium: A Puzzle", *Journal of Monetary Economics* 15, 145–161.

Merton, Robert C., 1969, "Lifetime Portfolio Selection Under Uncertainty: The Continuous Time Case", *Review of Economics and Statistics* 51, 247–257.

Merton, Robert C., 1971, "Optimum Consumption and Portfolio Rules in a Continuous-Time Model", *Journal of Economic Theory* 3, 373–413.

Merton, Robert C., 1973, "An Intertemporal Capital Asset Pricing Model", *Econometrica* 41, 867–87.

Merton, Robert C., 1990, *Continuous Time Finance*, Basil Blackwell, Cambridge, MA.

Modigliani, Franco and Richard Sutch, 1966, "Innovations in Interest Rate Policy", *American Economic Review* 56, 178–197.

Muirhead, Robb J., 1982, *Aspects of Multivariate Statistical Theory*, John Wiley & Sons, Inc., New York.

Nelson, Charles R. and Myung J. Kim, 1993, "Predictable Stock Returns: The Role of Small Sample Bias", *Journal of Finance* 48, 641–661.

Poterba, James M. and Lawrence H. Summers, 1988, "Mean Reversion in Stock Returns: Evidence and Implications", *Journal of Financial Economics* 22, 27–60.

Pratt, John W., 1964, "Risk-Aversion in the Small and in the Large", *Econometrica* 32, 122–136.

Restoy, Fernando, 1992, "Optimal Portfolio Policies Under Time-Dependent Returns", Bank of Spain Working Paper 9207, Bank of Spain, Madrid, Spain.

Samuelson, Paul A., 1969, "Lifetime Portfolio Selection by Dynamic Stochastic Programming", *Review of Economics and Statistics* 51, 239–246.

Samuelson, Paul A., 1979, "Why We Should Not Make Mean Log of Wealth Big Though Years to Act Are Long", *Journal of Banking and Finance* 3, 305–307.

Schroder, Mark and Costis Skiadas, 1999, "Optimal Consumption and Portfolio Selection with Stochastic Differential Utility", *Journal of Economic Theory* 89, 68–126.

Shiller, Robert J., 1979, "The Volatility of Long-Term Interest Rates and Expectations Models of the Term Structure", *Journal of Political Economy* 87, 1190–1219.

Shiller, Robert J., 1989, *Market Volatility*, MIT Press, Cambridge, MA.

Siegel, Jeremy, 1994, *Stocks for the Long Run*, McGraw-Hill, New York, NY.

Singleton, Kenneth J., 1990, "Specification and Estimation of Intertemporal Asset Pricing Models", in Ben Friedman and Frank Hahn eds. *Handbook of Monetary Economics*, Amsterdam: North-Holland.

Stambaugh, Robert F., 1999, "Predictive Regressions", *Journal of Financial Economics* 54, 375–421.

Teplá, Lucie, 1999, "Optimal Portfolio Policies with Borrowing and Short-sale Constraints", forthcoming *Journal of Economics Dynamics and Control*.

Tobin, James, 1958, "Liquidity Preference as Behavior Towards Risk", *Review of Economic Studies* 25, 68–85.

Vasicek, Oldrich, 1977, "An Equilibrium Characterization of the Term Structure", *Journal of Financial Economics* 5, 177–188.

Viceira, Luis M., 2001, "Optimal Portfolio Choice for Long-Horizon Investors with Nontradable Labor Income", forthcoming *Journal of Finance*.

Wachter, Jessica, 1998, "Risk Aversion and Allocation to Long-Term Bonds", unpublished paper, Harvard University.

Wachter, Jessica, 1999, "Portfolio and Consumption Decisions Under Mean-Reverting Returns: An Exact Solution", unpublished paper, Harvard University.

Weil, Philippe, 1989, "The Equity Premium Puzzle and the Risk-Free Rate Puzzle", *Journal of Monetary Economics* 24, 401–421.

Appendix to
Strategic Asset Allocation:
Portfolio Choice for Long-Term Investors

John Y. Campbell and Luis M. Viceira

July 2001

Contents

Chapter 1

Appendix: Introduction

This Appendix contains mathematical derivations of some selected results presented in John Campbell and Luis M. Viceira's book "Strategic Asset Allocation: Portfolio Choice for Long-Term Investors." To avoid confusion between equations in the main text of the book and equations in this Appendix, we number equations in the Appendix as (A1), (A2), etc.

Chapter 2

Appendix: Mathematical Derivations

2.1 Derivation of selected mathematical results in Chapter 3

2.1.1 Derivation of the approximation to the log portfolio return

In the case where there are two assets, one risky and one riskless, we have from (2.1) that

$$\frac{1+R_{p,t+1}}{1+R_{f,t+1}} = 1+\alpha_t\left(\frac{1+R_{t+1}}{1+R_{f,t+1}}-1\right).$$

Taking logs, this can be rewritten as

$$r_{p,t+1}-r_{f,t+1} = \log\left[1+\alpha_t\left(\exp(r_{t+1}-r_{f,t+1})-1\right)\right].$$

This equation gives a nonlinear relation between the log excess return on the single risky asset, $r_{t+1}-r_{f,t+1}$, and the log excess return on the portfolio, $r_{p,t+1}-r_{f,t+1}$. This relation can be approximated using a second-order Taylor expansion around the point $r_{t+1}-r_{f,t+1}=0$. The function $f_t(r_{t+1}-r_{f,t+1}) = \log\left[1+\alpha_t\left(\exp(r_{t+1}-r_{f,t+1})-1\right)\right]$ is approximated as

$$f_t(r_{t+1}-r_{f,t+1}) \approx f_t(0)+f_t'(0)(r_{t+1}-r_{f,t+1})+\frac{1}{2}f_t''(0)(r_{t+1}-r_{f,t+1})^2.$$

The derivatives of the function f_t, evaluated at $r_{t+1}-r_{f,t+1}=0$, are $f_t'(0) = \alpha_t$ and $f_t''(0)=\alpha_t(1-\alpha_t)$. Also, we replace $(r_{t+1}-r_{f,t+1})^2$ by its conditional expectation σ_t^2. Thus the Taylor approximation is

$$r_{p,t+1} - r_{f,t+1} = \alpha_t(r_{t+1} - r_{f,t+1}) + \frac{1}{2}\alpha_t(1-\alpha_t)\sigma_t^2.$$

The log excess portfolio return takes the same form as the simple excess portfolio return, with an adjustment factor in the variance of the risky asset return. The adjustment factor is zero if the portfolio weight in the risky asset is zero (for then the log portfolio return is just the log riskless return), and it is also zero if the weight in the risky asset is one (for then the log portfolio return is just the log risky return). The approximation in (2.21) can be justified rigorously by considering shorter and shorter time intervals. As the time interval shrinks, the higher-order terms that are dropped from (**??**) become negligible relative to those that are included, and the deviation of the realized squared excess return $(r_{t+1} - r_{f,t+1})^2$ from its expectation σ_t^2 also become negligible.

In the limit of continuous time, the approximation is exact and can be derived using Ito's Lemma. For completeness we present the derivation in the most general case where there are multiple risky assets and no riskless asset. The log return on the portfolio $r_{p,t+1}$ is a discrete-time approximation to its continuous-time counterpart. We assume that there are $(n+1)$ risky assets, one of which we use as a benchmark. Without loss of generality, we assume that the benchmark asset is a risky short-term instrument whose price we denote by B_t. We begin by specifying the return processes for the short-term instrument B_t and all other risky assets $\mathbf{P}_t$ in continuous time:

$$\begin{aligned} \frac{dB_t}{B_t} &= \mu_{b,t}dt + \boldsymbol{\sigma}_b d\mathbf{W}_t, \\ \frac{d\mathbf{P}_t}{\mathbf{P}_t} &= \boldsymbol{\mu}_t dt + \boldsymbol{\sigma} d\mathbf{W}_t, \end{aligned}$$

where $\mu_{b,t}$ and $\boldsymbol{\mu}_t$ are the drifts, $\boldsymbol{\sigma}_b$ and $\boldsymbol{\sigma}$ are the diffusion, and $\mathbf{W}_t$ is a m-dimensional standard Brownian motion. The dimensions of $\mu_b, \boldsymbol{\mu}, \boldsymbol{\sigma}_b, \boldsymbol{\sigma}$ are $1\times 1, n\times 1, 1\times m, n\times m$, respectively. We allow the drifts to depend on other state variables, but for notational simplicity, we suppress this dependency and simply use the time subscript. Moreover, note that the same $\mathbf{W}_t$ appears in these two equations.

Since we are working with log returns, we apply Ito's Lemma to each asset:

$$\begin{aligned} d\log B_t &= \left(\frac{dB_t}{B_t}\right) - \frac{1}{2}\left(\boldsymbol{\sigma}_b\boldsymbol{\sigma}_b'\right)dt, \\ d\log P_{i,t} &= \left(\frac{dP_{i,t}}{P_{i,t}}\right) - \frac{1}{2}\left(\boldsymbol{\sigma}_i\boldsymbol{\sigma}_i'\right)dt, \end{aligned}$$

where $\boldsymbol{\sigma}_i$ is the ith row of the diffusion matrix $\boldsymbol{\sigma}$, and $i = 1, .., n$.

Let V_t be the value of the portfolio at time t. We will use $d \log V_t$ to approximate $r_{p,t+1}$. By Ito's Lemma,

$$d \log V_t = \left(\frac{dV_t}{V_t}\right) - \frac{1}{2}\left(\frac{dV_t}{V_t}\right)^2.$$

We will now derive these 2 terms in order:

$$\begin{aligned}
\frac{dV_t}{V_t} &= \boldsymbol{\alpha}_t' \left(\frac{d\mathbf{P}_t}{\mathbf{P}_t}\right) + \left(1 - \boldsymbol{\alpha}_t'\boldsymbol{\iota}\right)\frac{dB_t}{B_t} \\
&= \boldsymbol{\alpha}_t' \left(d \log \mathbf{P}_t + \frac{1}{2}\left[\boldsymbol{\sigma}_i \boldsymbol{\sigma}_i'\right] dt\right) + \left(1 - \boldsymbol{\alpha}_t'\boldsymbol{\iota}\right)\left(d \log B_t + \frac{1}{2}\left(\boldsymbol{\sigma}_b \boldsymbol{\sigma}_b'\right) dt\right) \\
&= \boldsymbol{\alpha}_t' \left(d \log \mathbf{P}_t - d \log B_t \cdot \boldsymbol{\iota}\right) + d \log B_t \\
&\quad + \frac{1}{2}\boldsymbol{\alpha}_t' \left(\left[\boldsymbol{\sigma}_i \boldsymbol{\sigma}_i'\right] - \boldsymbol{\sigma}_b \boldsymbol{\sigma}_b' \cdot \boldsymbol{\iota}\right) dt + \frac{1}{2}\boldsymbol{\sigma}_b \boldsymbol{\sigma}_b' dt,
\end{aligned}$$

where $\boldsymbol{\iota}$ is a $n \times 1$ vector of ones and the bracket $[\cdot]$ denotes a vector with $\boldsymbol{\sigma}_i \boldsymbol{\sigma}_i'$ the ith entry. Next,

$$\begin{aligned}
\left(\frac{dV_t}{V_t}\right)^2 &= \boldsymbol{\alpha}_t' \left(d \log \mathbf{P}_t - d \log B_t \cdot \boldsymbol{\iota}\right)\left(d \log \mathbf{P}_t - d \log B_t \cdot \boldsymbol{\iota}\right)' \boldsymbol{\alpha}_t + \left(d \log B_t\right)^2 \\
&\quad + 2\boldsymbol{\alpha}_t' \left(d \log \mathbf{P}_t - d \log B_t \cdot \boldsymbol{\iota}\right)\left(d \log B_t\right) + o\left(dt\right),
\end{aligned}$$

where the $o(dt)$ terms vanish because they involve either $(dt)^2$ or $(dt)\left(d\mathbf{W}_t\right)$.

Now, from equation (**??**)–(**??**) and ignoring dt terms,

$$d \log \mathbf{P}_t - d \log B_t \cdot \boldsymbol{\iota} = \left(\boldsymbol{\sigma} - \boldsymbol{\iota} \cdot \boldsymbol{\sigma}_b\right) d\mathbf{W}_t.$$

Thus,

$$\begin{aligned}
\left(d \log \mathbf{P}_t - d \log B_t \cdot \boldsymbol{\iota}\right)\left(d \log \mathbf{P}_t - d \log B_t \cdot \boldsymbol{\iota}\right)' &= \left(\boldsymbol{\sigma} - \boldsymbol{\iota} \cdot \boldsymbol{\sigma}_b\right)\left(\boldsymbol{\sigma} - \boldsymbol{\iota} \cdot \boldsymbol{\sigma}_b\right)', \\
\left(d \log \mathbf{P}_t - d \log B_t \cdot \boldsymbol{\iota}\right)\left(d \log B_t\right) &= \left(\boldsymbol{\sigma} - \boldsymbol{\iota} \cdot \boldsymbol{\sigma}_b\right) \cdot \boldsymbol{\sigma}_b'.
\end{aligned}$$

Collecting these results and using our notation for excess returns and the return on the benchmark risky asset ($\mathbf{x}_{t+1} = d \log \mathbf{P}_t - d \log B_t \cdot \boldsymbol{\iota}$, and

$r_{0,t+1} = d\log(B_t))$ and setting $dt = 1$, we have:

$$\begin{aligned}
& r_{p,t+1} \\
= \; & d\log V_t \\
= \; & \boldsymbol{\alpha}_t' \mathbf{x}_{t+1} + r_{0,t+1} + \frac{1}{2}\boldsymbol{\alpha}_t' \left(\left[\boldsymbol{\sigma}_i \boldsymbol{\sigma}_i' \right] - \boldsymbol{\sigma}_b \boldsymbol{\sigma}_b' \cdot \boldsymbol{\iota} \right) \\
& -\frac{1}{2} \left[\boldsymbol{\alpha}_t' \left(\boldsymbol{\sigma} - \boldsymbol{\iota} \cdot \boldsymbol{\sigma}_b \right) \left(\boldsymbol{\sigma} - \boldsymbol{\iota} \cdot \boldsymbol{\sigma}_b \right)' \boldsymbol{\alpha}_t + 2\boldsymbol{\alpha}_t' \left(\boldsymbol{\sigma} - \boldsymbol{\iota} \cdot \boldsymbol{\sigma}_b \right) \boldsymbol{\sigma}_b' \right] \\
= \; & \boldsymbol{\alpha}_t' \mathbf{x}_{t+1} + r_{0,t+1} - \frac{1}{2}\boldsymbol{\alpha}_t' \left(\boldsymbol{\sigma} - \boldsymbol{\iota} \cdot \boldsymbol{\sigma}_b \right) \left(\boldsymbol{\sigma} - \boldsymbol{\iota} \cdot \boldsymbol{\sigma}_b \right)' \boldsymbol{\alpha}_t \\
& +\frac{1}{2}\boldsymbol{\alpha}_t' \left(\left[\boldsymbol{\sigma}_i \boldsymbol{\sigma}_i' \right] + \boldsymbol{\sigma}_b \boldsymbol{\sigma}_b' \cdot \boldsymbol{\iota} - \mathbf{2} \boldsymbol{\sigma} \boldsymbol{\sigma}_b' \right).
\end{aligned}$$

Similarly, using the notation in book for variances and covariances, we have

$$\begin{aligned}
\left(\boldsymbol{\sigma} - \boldsymbol{\iota} \cdot \boldsymbol{\sigma}_b \right) \left(\boldsymbol{\sigma} - \boldsymbol{\iota} \cdot \boldsymbol{\sigma}_b \right)' &\equiv \Sigma_t, \\
\boldsymbol{\sigma}_b \boldsymbol{\sigma}_b' &\equiv \sigma_{0t}^2, \\
\left[\boldsymbol{\sigma}_i \boldsymbol{\sigma}_i' \right] + \boldsymbol{\sigma}_b \boldsymbol{\sigma}_b' \cdot \boldsymbol{\iota} - \mathbf{2} \boldsymbol{\sigma} \boldsymbol{\sigma}_b' &= \boldsymbol{\sigma}_t^2.
\end{aligned}$$

With these terms, the return on the portfolio is

$$r_{p,t+1} = \boldsymbol{\alpha}_t' \mathbf{x}_{t+1} + r_{0,t+1} + \frac{1}{2}\boldsymbol{\alpha}_t' \boldsymbol{\sigma}_t^2 - \frac{1}{2}\boldsymbol{\alpha}_t' \Sigma_t \boldsymbol{\alpha}_t,$$

which is equation (2.23) in text.

2.2 Derivation of selected mathematical results in Chapter 3

2.2.1 Derivation of the approximation to the log intertemporal budget constraint

Taking logs on both sides of the intertemporal budget constraint (3.2) we obtain equation (3.3) in text:

$$\Delta w_{t+1} = r_{p,t+1} + \log(1 - \exp(c_t - w_t)). \tag{A1}$$

The second-term on the right-hand side of (A1) is a non-linear function of the log consumption-wealth ratio. A first-order approximation of this function around the mean of the log consumption-wealth ratio gives:

$$\begin{aligned}\log(1-\exp(c_t-w_t)) &\approx \log\left(1-\exp\left(\mathrm{E}\left[c_t-w_t\right]\right)\right)\\ &\quad -\frac{\exp\left(\mathrm{E}\left[c_t-w_t\right]\right)}{1-\exp\left(\mathrm{E}\left[c_t-w_t\right]\right)}\left(\left(c_t-w_t\right)-\mathrm{E}\left[c_t-w_t\right]\right)\\ &= \log\left(1-\exp\left(\mathrm{E}\left[c_t-w_t\right]\right)\right)\\ &\quad +\frac{\exp\left(\mathrm{E}\left[c_t-w_t\right]\right)}{1-\exp\left(\mathrm{E}\left[c_t-w_t\right]\right)}\mathrm{E}\left[c_t-w_t\right]\\ &\quad -\frac{\exp\left(\mathrm{E}\left[c_t-w_t\right]\right)}{1-\exp\left(\mathrm{E}\left[c_t-w_t\right]\right)}\left(c_t-w_t\right). \end{aligned} \tag{A2}$$

Defining

$$\rho \equiv 1-\exp\left(\mathrm{E}\left[c_t-w_t\right]\right), \tag{A3}$$

we can rewrite (A2) as

$$\log(1-\exp(c_t-w_t)) \approx k+\left(1-\frac{1}{\rho}\right)\left(c_t-w_t\right), \tag{A4}$$

where

$$\begin{aligned} k &\equiv \log\left(1-\exp\left(\mathrm{E}\left[c_t-w_t\right]\right)\right)+\frac{\exp\left(\mathrm{E}\left[c_t-w_t\right]\right)}{1-\exp\left(\mathrm{E}\left[c_t-w_t\right]\right)}\left(\mathrm{E}\left[c_t-w_t\right]\right)\\ &= \log\left(\rho\right)+\frac{1-\rho}{\rho}\log\left(1-\rho\right). \end{aligned}$$

Note that this approximation is exact when the optimal consumption-wealth ratio is constant. so that $c_t-w_t=\mathrm{E}_t[c_t-w_t]$.

Direct substitution of (A4) into (A1) gives equation (3.4) for the log intertemporal budget constraint in text.

2.2.2 Solution to model with constant variances and risk premia when there are multiple risky assets

Chapter 2 shows (see equation [2.51]) that under Epstein-Zin utility with multiple risky assets, the premium on each risky asset over the risky benchmark asset is given by

$$\begin{aligned} &\mathrm{E}_t\left(r_{i,t+1}-r_{0,t+1}\right)+\frac{1}{2}\mathrm{Var}_t\left(r_{i,t+1}-r_{0,t+1}\right)\\ = &\ \frac{\theta}{\psi}\mathrm{Cov}_t\left(\Delta c_{t+1},r_{i,t+1}-r_{0,t+1}\right)\\ &+\left(1-\theta\right)\mathrm{Cov}_t\left(r_{p,t+1},r_{i,t+1}-r_{0,t+1}\right)\\ &-\mathrm{Cov}_t\left(r_{i,t+1}-r_{0,t+1},r_{0,t+1}\right), \end{aligned} \tag{A5}$$

where $\theta = (1-\gamma)/(1-\psi^{-1})$.

Using the log budget constraint (3.4) and the trivial identity $\Delta c_{t+1} = (c_{t+1} - w_{t+1}) - (c_t - w_t) + \Delta w_{t+1}$, we have

$$\begin{aligned} & \mathrm{Cov}_t\left(\Delta c_{t+1}, r_{i,t+1} - r_{0,t+1}\right) \\ = \; & \mathrm{Cov}_t\left((c_{t+1} - w_{t+1}) - (c_t - w_t) + \Delta w_{t+1}, r_{i,t+1} - r_{0,t+1}\right) \\ = \; & \mathrm{Cov}_t\left(c_{t+1} - w_{t+1}, r_{i,t+1} - r_{0,t+1}\right) + \mathrm{Cov}_t\left(\Delta w_{t+1}, r_{i,t+1} - r_{0,t+1}\right) \\ = \; & \mathrm{Cov}_t\left(c_{t+1} - w_{t+1}, r_{i,t+1} - r_{0,t+1}\right) + \mathrm{Cov}_t\left(r_{pt+1}, r_{i,t+1} - r_{0,t+1}\right). \end{aligned} \tag{A6}$$

Substitution of (A6) into (A5) gives

$$\begin{aligned} & \mathrm{E}_t\left(r_{i,t+1} - r_{0,t+1}\right) + \frac{1}{2}\mathrm{Var}_t\left(r_{i,t+1} - r_{0,t+1}\right) \\ = \; & \frac{\theta}{\psi}\mathrm{Cov}_t\left(c_{t+1} - w_{t+1}, r_{i,t+1} - r_{0,t+1}\right) \\ & + \left(1 - \theta + \frac{\theta}{\psi}\right)\mathrm{Cov}_t\left(r_{p,t+1}, r_{i,t+1} - r_{0,t+1}\right) \\ & - \mathrm{Cov}_t\left(r_{i,t+1} - r_{0,t+1}, r_{0,t+1}\right), \end{aligned}$$

or, in vector notation,

$$\begin{aligned} & \mathrm{E}_t\left(\mathbf{r}_{t+1} - r_{0,t+1}\boldsymbol{\iota}\right) + \frac{1}{2}\boldsymbol{\sigma}_t^2 \\ = \; & -\left(\frac{1-\gamma}{1-\psi}\right)\mathrm{Cov}_t\left(c_{t+1} - w_{t+1}, \mathbf{r}_{t+1} - r_{0,t+1}\boldsymbol{\iota}\right) \\ & + \gamma\,\mathrm{Cov}_t\left(r_{p,t+1}, \mathbf{r}_{t+1} - r_{0,t+1}\boldsymbol{\iota}\right) \\ & - \mathrm{Cov}_t\left(\mathbf{r}_{t+1} - r_{0,t+1}\boldsymbol{\iota}, r_{0,t+1}\right), \end{aligned} \tag{A7}$$

where we have substituted $(1-\gamma)/(1-\psi^{-1})$ for θ. $\boldsymbol{\sigma}_t^2$ denotes a column vector with the variance of the excess return on each asset over the return on the benchmark risky asset:

$$\boldsymbol{\sigma}_t^2 \equiv \left(\mathrm{Var}_t\left(r_{1,t+1} - r_{0,t+1}\right), ..., \mathrm{Var}_t\left(r_{n,t+1} - r_{0,t+1}\right)\right)'.$$

The equation for log portfolio return (2.23) implies that the second covariance term in (A7) is

$$\begin{aligned} & \mathrm{Cov}_t\left(r_{p,t+1}, \mathbf{r}_{t+1} - r_{0,t+1}\boldsymbol{\iota}\right) \\ = \; & \mathrm{Cov}_t\left((\mathbf{r}_{t+1} - r_{0,t+1}\boldsymbol{\iota}) + r_{0,t+1}, \mathbf{r}_{t+1} - r_{0,t+1}\boldsymbol{\iota}\right) \\ = \; & \boldsymbol{\alpha}_t'\,\mathrm{Var}_t\left(\mathbf{r}_{t+1} - r_{0,t+1}\boldsymbol{\iota}\right) + \mathrm{Cov}_t\left(r_{0,t+1}, \mathbf{r}_{t+1} - r_{0,t+1}\boldsymbol{\iota}\right), \end{aligned}$$

so that (A7) becomes

$$\begin{aligned}
& \mathrm{E}_t\left(\mathbf{r}_{t+1}-r_{0,t+1}\boldsymbol{\iota}\right)+\frac{1}{2}\boldsymbol{\sigma}_t^2 \\
= & -\left(\frac{1-\gamma}{1-\psi}\right)\mathrm{Cov}_t\left(c_{t+1}-w_{t+1},\mathbf{r}_{t+1}-r_{0,t+1}\boldsymbol{\iota}\right) \\
& +\gamma\boldsymbol{\alpha}_t'\,\mathrm{Var}_t\left(\mathbf{r}_{t+1}-r_{0,t+1}\boldsymbol{\iota}\right) \\
& -(1-\gamma)\,\mathrm{Cov}_t\left(\mathbf{r}_{t+1}-r_{0,t+1}\boldsymbol{\iota},r_{0,t+1}\right), \qquad \text{(A8)}
\end{aligned}$$

from which we obtain immediately an expression for $\boldsymbol{\alpha}_t$.

Defining

$$\boldsymbol{\sigma}_{ht}\equiv\frac{\mathrm{Cov}_t\left(-\left(c_{t+1}-w_{t+1}\right),\mathbf{r}_{t+1}-r_{0,t+1}\boldsymbol{\iota}\right)}{1-\psi},$$

$$\boldsymbol{\Sigma}_t\equiv\mathrm{Var}_t\left(\mathbf{r}_{t+1}-r_{0,t+1}\boldsymbol{\iota}\right),$$

and

$$\boldsymbol{\sigma}_{0t}\equiv\mathrm{Cov}_t\left(\mathbf{r}_{t+1}-r_{0,t+1}\boldsymbol{\iota},r_{0,t+1}\right),$$

the expression for $\boldsymbol{\alpha}_t$ resulting from (A8) becomes

$$\boldsymbol{\alpha}_t=\frac{1}{\gamma}\boldsymbol{\Sigma}_t^{-1}\left(\mathrm{E}_t\,\mathbf{r}_{t+1}-r_{0,t+1}\boldsymbol{\iota}+\boldsymbol{\sigma}_t^2/2\right)+\left(1-\frac{1}{\gamma}\right)\boldsymbol{\Sigma}_t^{-1}\left(\boldsymbol{\sigma}_{ht}-\boldsymbol{\sigma}_{0t}\right),$$

which is (3.21) in text. Equation (3.20) obtains when the benchmark asset is riskless one-period ahead, so that $\boldsymbol{\sigma}_{0t}=0$.

Note that section 3.1.3 shows that

$$\frac{\left(\mathrm{E}_{t+1}-\mathrm{E}_t\right)\left(c_{t+1}-w_{t+1}\right)}{1-\psi}=\left(\mathrm{E}_{t+1}-\mathrm{E}_t\right)\sum_{j=1}^{\infty}\rho^j r_{p,t+1+j}. \qquad \text{(A9)}$$

This section also shows that, when time-variation in interest rates is the only source of variation in investment opportunities, the right-hand-side of (A9) is equal to

$$\left(\mathrm{E}_{t+1}-\mathrm{E}_t\right)\sum_{j=1}^{\infty}\rho^j r_{p,t+1+j}=\left(\mathrm{E}_{t+1}-\mathrm{E}_t\right)\sum_{j=1}^{\infty}\rho^j r_{f,t+1+j},$$

so that

$$\begin{aligned}
\boldsymbol{\sigma}_{ht} & \equiv \mathrm{Cov}_t\left(-\frac{c_{t+1}-w_{t+1}}{1-\psi},\mathbf{r}_{t+1}-r_{0,t+1}\boldsymbol{\iota}\right) \\
& = \mathrm{Cov}_t\left(-\left(\mathrm{E}_{t+1}-\mathrm{E}_t\right)\sum_{j=1}^{\infty}\rho^j r_{f,t+1+j},\mathbf{r}_{t+1}-r_{0,t+1}\boldsymbol{\iota}\right), \qquad \text{(A10)}
\end{aligned}$$

as stated in (3.19).

From equation (3.18) in text, it is easy to see that, if $\rho = \rho_c$, then equation (A10) becomes

$$\begin{aligned}\boldsymbol{\sigma}_{ht} &\equiv \mathrm{Cov}_t\left(-(\mathrm{E}_{t+1}-\mathrm{E}_t)\sum_{j=1}^{\infty}\rho^j r_{f,t+1+j}, \mathbf{r}_{t+1}-r_{0,t+1}\boldsymbol{\iota}\right)\\ &= \mathrm{Cov}_t\left(r_{c,t+1}, \mathbf{r}_{t+1}-r_{0,t+1}\boldsymbol{\iota}\right)\\ &\equiv \boldsymbol{\sigma}_{ct},\end{aligned}$$

so that $\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\sigma}_{ht} = \boldsymbol{\Sigma}_t^{-1}\boldsymbol{\sigma}_{ct}$ is the vector of population regression coefficients from a multiple regression of an inflation-indexed consol return onto the set of risky asset returns, as stated in text.

2.2.3 Recursive expression for A_n

The recursive equation for the coefficient A_n in the indexed zero-coupon bond pricing equation (3.27) is given by

$$A_n - A_{n-1} = (1-\phi_x)\,\mu_x B_{n-1} - \frac{1}{2}\left[(\beta_{mx}+B_{n-1})^2\,\sigma_x^2 + \sigma_m^2\right].$$

with $A_0 = B_0 = 0$. See Campbell, Lo and Mackinlay (1997) for a derivation of the pricing equation (3.27).

2.2.4 Pricing nominal bonds

The pricing of default-free nominal bonds follows the same steps as the pricing of indexed bonds. The relevant stochastic discount factor to price nominal bonds is the nominal SDF $M_{t+1}^{\$}$, whose log is given

$$m_{t+1}^{\$} = m_{t+1} - \pi_{t+1}. \tag{A11}$$

Since both M_{t+1} and Π_{t+1} are jointly lognormal and homoskedastic, $M_{t+1}^{\$}$ is also lognormal. The log nominal return on a one-period nominal bond is $r_{1,t+1}^{\$} = -\log \mathrm{E}_t[M_{t+1}]$, or

$$\begin{aligned}r_{1,t+1}^{\$} &= -\mathrm{E}_t\left[m_{t+1}^{\$}\right] - \frac{1}{2}\mathrm{Var}_t\left[m_{t+1}^{\$}\right]\\ &= x_t + z_t - \frac{1}{2}\left[(\beta_{mx}+\beta_{\pi x})\,2\sigma_x^2 + \beta_{\pi z}^2\sigma_z^2 + (1+\beta_\pi)^2\,\sigma_m^2 + \sigma_\pi^2\right],\end{aligned}$$

a linear combination of the expected log real SDF and expected inflation.

The risk premium on a 1-period nominal bond over a 1-period real bond can be written as

$$\mathrm{E}_t\left[r^{\$}_{1,t+1} - \pi_{t+1} - r_{1,t+1}\right] + \frac{1}{2}\,\mathrm{Var}_t\left[\pi_{t+1}\right] = -\beta_{mx}\beta_{\pi x}\sigma_x^2 - \beta_{\pi m}\sigma_m^2,$$

which has the same form as equation (3.38) for equities.

The log price of an n-period nominal bond, $p^{\$}_{n,t}$, also has an affine structure. It is a linear combination of x_t and z_t whose coefficients are time-invariant, though they vary with the maturity of the bond. As shown in equation (3.36), $-p^{\$}_{n,t} = A^{\$}_n + B^{\$}_{1,n}x_t + B^{\$}_{2,n}z_t$,where

$$\begin{aligned}
B^{\$}_{1,n} &= 1 + \phi_x B^{\$}_{1,n-1} = \frac{1-\phi_x^n}{1-\phi_x} \\
B^{\$}_{2,n} &= 1 + \phi_z B^{\$}_{2,n-1} = \frac{1-\phi_z^n}{1-\phi_z} \\
A^{\$}_n - A^{\$}_{n-1} &= (1-\phi_x)\,\mu_x B^{\$}_{1,n-1} + (1-\phi_z)\,\mu_z B^{\$}_{2,n-1} \\
&\quad -\frac{1}{2}\left(\beta_{mx} + \beta_{\pi_x} + B^{\$}_{1,n-1} + \beta_{zx}B^{\$}_{2,n-1}\right)^2 \sigma_x^2 \\
&\quad -\frac{1}{2}\left(\beta_{\pi z} + B^{\$}_{2,n-1}\right)^2 \sigma_z^2 - \frac{1}{2}\left(1 + \beta_{\pi m} + \beta_{zm}\right)^2 \sigma_m^2 \\
&\quad -\frac{1}{2}\sigma_\pi^2,
\end{aligned}$$

and $A^{\$}_0 = B^{\$}_{1,0} = B^{\$}_{2,0} = 0$.

The excess return on a n-period bond over the one-period log nominal interest rate is

$$\begin{aligned}
r^{\$}_{n,t+1} - r^{\$}_{1,t+1} &= p^{\$}_{n-1,t+1} - p^{\$}_{n,t} + p^{\$}_{1,t} \\
&= -\left(B^{\$}_{1,n-1} + B^{\$}_{2,n-1}\beta_{zx}\right)\left(\beta_{mx} + \beta_{\pi x}\right)\sigma_x^2 - B^{\$}_{2,n-1}\beta_\pi\sigma_z^2 \\
&\quad -\left(1 + \beta_{\pi m}\right)\beta_{zm}\sigma_m^2 - \frac{1}{2}\left(B^{\$}_{1,n-1} + B^{\$}_{2,n-1}\beta_{zx}\right)2\sigma_x^2 \\
&\quad -\frac{1}{2}\left(B^{\$}_{2,n-1}\right)^2\sigma_z^2 - \frac{1}{2}\beta_{zm}^2\sigma_m^2 \\
&\quad -\left(B^{\$}_{1,n-1} + B^{\$}_{2,n-1}\beta_{zx}\right)\varepsilon_{x,t+1} - B^{\$}_{2,n-1}\beta_{zm}\varepsilon_{m,t+1} \\
&\quad -B^{\$}_{2,n-1}\varepsilon_{z,t+1}.
\end{aligned}$$

The terms in $B^{\$}_{2,n-1}\beta_{zx}$ and $B^{\$}_{2,n-1}\beta_{zm}$ arise because shocks to expected inflation are correlated with shocks to the expected and unexpected log real SDF. Thus risk premia in the nominal term structure are different from risk premia in the real term structure because they include compensation for inflation risk. Like real risk premia, however, nominal risk premia are constant over time.

2.3 Derivation of selected mathematical results in Chapter 5

2.3.1 Coefficients of the value function in the model with time-varying interest rates and power utility

Substitution of (5.22) into (5.20) leads to

$$\begin{aligned} 0 &= \frac{\gamma h_0}{1-\gamma} - \frac{\gamma h_1}{1-\gamma}\left(C_0 + C_1 r\right) + \frac{\lambda^2}{2\gamma\sigma^2} - \frac{\beta}{1-\gamma} + r \\ &\quad + \left(\frac{\gamma\kappa}{1-\gamma}\left(\theta - r\right) - \lambda\right) C_1 + \frac{\gamma\sigma^2}{2\left(1-\gamma\right)} C_1^2, \end{aligned} \tag{A12}$$

where we must determine C_0 and C_1 so that the equation holds for all values of the instantaneous interest rate. Simple inspection of the terms in the equation shows that the right hand side of the equation is a linear combination of the instantaneous interest rate. Thus C_0 and C_1 must be such that both the intercept and the slope of the linear equation are zero simultaneously:

$$0 = -\frac{\gamma\left(h_1 + \kappa\right)}{1-\gamma} C_1 + 1, \tag{A13}$$

$$\begin{aligned} 0 &= -\frac{\gamma h_1}{1-\gamma} C_0 + \frac{\gamma h_0}{1-\gamma} + \frac{\lambda^2}{2\gamma\sigma^2} - \frac{\beta}{1-\gamma} \\ &\quad \left(\frac{\gamma\kappa\theta}{1-\gamma} - \lambda\right) C_1 + \frac{\gamma\sigma^2}{2\left(1-\gamma\right)} C_1^2. \end{aligned} \tag{A14}$$

Equation (A13) is a linear equation whose only unknown is C_1. The solution to this equation is given in (5.23). Equation (A14) depends on both C_1 and C_0, but it is linear in C_0 given C_1. Substituting the expression for C_1 that obtains from (A13) into (A14), we can solve for C_0 immediately.

2.3.2 Coefficients of the value function in the model with time-varying interest rates and recursive utility (unit elasticity of intertemporal substitution case)

The solution procedure is analogous to the solution procedure shown in the previous section. Substitution of (5.72) into (5.71) leads to

$$\begin{aligned} 0 &= -\frac{\beta}{1-\gamma}\left(C_0+C_1 r\right)+\left(\beta\log\beta-\beta+\frac{\lambda^2}{2\gamma\sigma^2}+r\right) \\ &\quad +\frac{\sigma^2}{2\gamma}C_1^2+\left(\frac{\kappa}{1-\gamma}\left(\theta-r\right)-\frac{\lambda}{\gamma}\right)C_1+\frac{\sigma^2}{2\left(1-\gamma\right)}C_1^2, \qquad \text{(A15)} \end{aligned}$$

where we must determine C_0 and C_1 so that the equation holds for all values of the instantaneous interest rate. Once again, simple inspection of the terms in the equation shows that the right hand side of the equation is a linear combination of the instantaneous interest rate. Thus C_0 and C_1 must be such that both the intercept and the slope of the linear equation are zero simultaneously. This leads to two algebraic equations for C_0 and C_1.

The first equation obtains from collecting terms in r in (A15) and setting them to zero:

$$0=-\frac{\gamma\left(\beta+\kappa\right)}{1-\gamma}C_1+1. \qquad \text{(A16)}$$

This equation is identical to (A13) with $h_1=\beta$. The second equation is again linear in C_0 given C_1, and obtains by collecting all other terms in (A15).

2.3.3 Coefficients of the value function in the model with stochastic volatility and recursive utility (unit elasticity of intertemporal substitution case)

Substitution of guess $I=\exp\{C_0+C_1 q_t\}$ into (5.85) leads to

$$\begin{aligned} 0 &= \beta\log\beta-\beta-\frac{\beta\gamma}{1-\gamma}\log\left(1-\gamma\right)+\frac{\left(\mu_P-r\right)^2}{2\gamma}q+r \\ &\quad -\frac{1}{1-\gamma}\beta\left(C_0+C_1 q\right)+\left(\frac{\rho_{Pq}\sigma_q\left(\mu_P-r\right)}{\gamma}y+\frac{\kappa_q}{1-\gamma}\left(\theta_q-q\right)\right)C_1 \\ &\quad +\frac{\rho_{Pq}^2\sigma_q^2}{2\gamma}qC_1^2+\frac{\sigma_q^2}{2\left(1-\gamma\right)}qC_1^2. \qquad \text{(A17)} \end{aligned}$$

where we must determine C_0 and C_1 so that the equation holds for all values of precision q_t. Once again, simple inspection of the terms in the equation shows that the right hand side of the equation is a linear combination of q_t. Thus C_0 and C_1 must be such that both the intercept and the slope of the linear equation are zero simultaneously. This leads to the following two algebraic equations for C_0 and C_1:

$$0 = aC_1^2 + bC_1 + c, \tag{A18}$$
$$0 = (1-\gamma)(\beta \log \beta + r - \beta) - \beta\gamma \log(1-\gamma) - \beta C_0 + \kappa_q \theta_q C_1, \tag{A19}$$

where

$$a = \frac{\sigma_q^2}{2\gamma(1-\gamma)}\left[\gamma\left(1-\rho_{Pq}^2\right) + \rho_{Pq}^2\right], \tag{A20}$$
$$b = \frac{\rho_{Pq}\sigma_q(\mu_P - r)}{\gamma} - \frac{\beta+\kappa_q}{1-\gamma}, \tag{A21}$$
$$c = \frac{(\mu_P - r)^2}{2\gamma}. \tag{A22}$$

Equation(A18) is a quadratic equation in C_1, and equation (A19) is linear in C_0 given C_1. For general parameter values the equation for C_1 has two roots. These roots are always real provided that $\gamma > 1$. From standard theory on quadratic equations, the product of the roots is equal to c/a. When $\gamma > 1$, this ratio is always negative so that the roots have opposite signs. It is easy to check that only the negative root maximizes the value function for all values of q_t.[1] This root is obtained by selecting the positive root of the discriminant of the quadratic equation. Therefore, $C_1 < 0$ when $\gamma > 1$.

When $\gamma < 1$, the roots are real—and a solution to the problem exists—if and only if

$$\left(\frac{1-\gamma}{\gamma}\frac{\sigma_q(\mu_P - r)}{\beta+\kappa_q}\right)\left(2\rho_{Pq} + \frac{\sigma_q(\mu_P - r)}{\beta+\kappa_q}\right) < 1.$$

This condition implies that both roots of the quadratic equation are positive. In this case the largest root—again, the root associated with the positive root of the discriminant—maximizes the value function. Therefore, $C_1 > 0$ when $\gamma < 1$. Putting together the results for $\gamma > 1$ and $\gamma < 1$, we have that $\widetilde{C}_1 = C_1/(1-\gamma) > 0$.

[1] Note that the equation for B implies that $\partial C_0/\partial C_1 > 0$.

2.4 Derivation of selected mathematical results in Chapter 6

2.4.1 Optimal consumption and portfolio choice for retired investors

Optimal portfolio rule

A retired investor does not have labor income. Thus he faces the intertemporal budget constraint

$$W_{t+1}^r = (W_t - C_t^r)\left(1 + R_{p,t+1}^r\right),$$

whose loglinear approximation is given in equation (3.4) in text. To facilitate comparisons with the labor income case, it is convenient to rewrite (3.4) as follows:

$$w_{t+1}^r - w_t = k^r - \rho_c^r\,(c_t^r - w_t) + r_{p,t+1}^r. \tag{A23}$$

where $\rho_c^r \equiv -(1 - 1/\rho) = \exp\{\mathrm{E}[c^r - w_t]\}/(1 - \exp\{\mathrm{E}[c^r - w_t]\})$, and $k^r = -(1 + \rho_c^r)\log(1 + \rho_c^r) + \rho_c^r \log(\rho_c^r)$. Note that (A23) holds exactly when the consumption-wealth ratio is constant—as it is in this case.

We have also shown in this appendix that we can approximate the log portfolio return with the following expression:

$$r_{p,t+1} = r_f + \alpha_t(r_{t+1} - r_f) + \frac{1}{2}\alpha_t(1 - \alpha_t)\sigma_u^2. \tag{A24}$$

This is equation (2.21) when investment opportunities are constant.

We have shown in section 2.2.3 that the Euler equation for an investor with power utility of consumption and no labor income implies the following expression for the risk-premium on the risky asset (see equation [2.43]):

$$\mathrm{E}_t\, r_{t+1} - r_f + \frac{1}{2}\,\mathrm{Var}_t\,(r_{t+1}) = \gamma\,\mathrm{Cov}_t\left(c_{t+1}^r - c_t^r, r_{t+1}\right). \tag{A25}$$

or, given our assumptions about the investment opportunity set,

$$\mu + \frac{1}{2}\sigma_u^2 = \gamma\,\mathrm{Cov}_t\left(c_{t+1}^r - c_t^r, r_{t+1}\right). \tag{A26}$$

We can compute the covariance term in the right-hand side of equation (A26) by noting that (6.42) implies

$$c_{t+1}^r - c_t^r = b_1^r(w_{t+1}^r - w_t), \tag{A27}$$

so that

$$\begin{aligned}\mathrm{Cov}_t\left(c^r_{t+1}-c^r_t, r_{t+1}\right) &= b^r_1\,\mathrm{Cov}_t\left(w^r_{t+1}-w_t, r_{t+1}\right)\\ &= b^r_1\,\mathrm{Cov}_t\left(r_{p,t+1}, r_{t+1}\right)\\ &= b^r_1\alpha^r_t\sigma^2_u,\end{aligned}$$

where the second equality obtains from (A23), and the third equality obtains from (A24).

Therefore,

$$\mu+\frac{1}{2}\sigma^2_u=\gamma b^r_1\alpha^r_t\sigma^2_u,$$

from which (6.43) in text follows.

Optimal consumption rule

To derive the optimal consumption rule, note that the log of the Euler equation (6.41) with $i = p$ yields the following equation for expected log consumption growth:

$$\mathrm{E}_t\left[c^r_{t+1}-c^r_t\right]=\frac{1}{\gamma}\left[\mathrm{E}\left[r^r_{p,t+1}\right]+\log\delta^r+\frac{1}{2}\mathrm{Var}\left[r^r_{p,t+1}-\gamma\left(c^r_{t+1}-c^r_t\right)\right]\right], \tag{A28}$$

where $\delta^r \equiv (1-\pi^d)\delta$, and I have suppressed the subscript t from the conditional moments on the right-hand-side because $\alpha^r_t \equiv \alpha^r$ implies that they are constant. Moreover, equation (A27) implies $\mathrm{Var}_t[r^r_{p,t+1}-\gamma(c^r_{t+1}-c^r_t)] = (1-b^r_1\gamma)^2\,\mathrm{Var}(r^r_{p,t+1})$.

On the other hand, from equation (A27) and the log budget constraint (A23) we have

$$\begin{aligned}\mathrm{E}_t\left[c^r_{t+1}-c^r_t\right] &= b^r_1\,\mathrm{E}_t\left[w^r_{t+1}-w_t\right]\\ &= b^r_1\,\mathrm{E}\left[r^r_{p,t+1}\right]-b^r_1\rho^r_c b^r_0+b^r_1k^r+b^r_1\rho^r_c\left(1-b^r_1\right)w_t.\end{aligned} \tag{A29}$$

Equalizing the right-hand side of equations (A28) and (A29), and identifying coefficients, we obtain two equations. The first one implies

$$b^r_1=1,$$

while the second one implies

$$\begin{aligned}b^r_0 = &-\left(\frac{1}{b^r_1\rho^r_c}\right)\left[\left(\frac{1}{\gamma}-b^r_1\right)\mathrm{E}\left[r^r_{p,t+1}\right]+\frac{1}{\gamma}\log\delta^r\right.\\ &\left.+\frac{1}{2\gamma}\left(1-b^r_1\gamma\right)^2\mathrm{Var}\left(r^r_{p,t+1}\right)-b^r_1k^r\right].\end{aligned} \tag{A30}$$

Since $b_1^r = 1$, we have $b_0^r \equiv \mathrm{E}[c_t^r - w_t] = \log(\rho_c^r) - \log(1+\rho_c^r)$. We can easily substitute the log-linearization constants ρ_c^r and k^r out from equation (A30) and obtain

$$b_0^r = \log\left(1 - \exp\left\{\left[\left(\frac{1}{\gamma} - b_1^r\right)\mathrm{E}\left[r_{p,t+1}^r\right] + \frac{1}{\gamma}\log\delta^r + \frac{1}{2\gamma}\left(1 - b_1^r\gamma\right)^2 \mathrm{Var}\left(r_{p,t+1}^r\right)\right]\right\}\right). \tag{A31}$$

2.4.2 Optimal consumption and portfolio choice for employed investors

To derive the optimal portfolio rule in the employment state we first need to derive loglinear expressions for the intertemporal budget constraint (6.39) and the Euler equation (6.40) for an employed investor.

Loglinear intertemporal budget constraint

We can rewrite the intertemporal budget constraint (6.39) as

$$\frac{W_{t+1}^e}{L_{t+1}} = \left(1 + \frac{W_t}{L_t} - \frac{C_t^e}{L_t}\right)\left(\frac{L_t}{L_{t+1}}\right) R_{p,t+1}^e,$$

or, in logs,

$$w_{t+1}^e - l_{t+1} = \log\left(\exp\left\{w_t - l_t\right\} - \exp\left\{c_t^e - l_t\right\}\right) - \Delta l_{t+1} + r_{p,t+1}^e. \tag{A32}$$

We can now linearize equation (A32) by taking a first-order Taylor expansion around $(c_t^e - l_t) = \mathrm{E}[c_t^e - l_t]$ and $(w_t^e - l_t) = \mathrm{E}[w_t^e - l_t]$. This gives

$$w_{t+1}^e - l_{t+1} \approx k^e + \rho_w^e\left(w_t - l_t\right) - \rho_c^e\left(c_t^e - l_t\right) - \Delta l_{t+1} + r_{p,t+1}^e, \tag{A33}$$

where

$$\rho_w^e = \frac{\exp\left\{\mathrm{E}[w_t^e - l_t]\right\}}{1 + \exp\left\{\mathrm{E}[w_t^e - l_t]\right\} - \exp\left\{\mathrm{E}[c_t^e - l_t]\right\}}, \tag{A34}$$

$$\rho_c^e = \frac{\exp\left\{\mathrm{E}[c_t^e - l_t]\right\}}{1 + \exp\left\{\mathrm{E}[w_t^e - l_t]\right\} - \exp\left\{\mathrm{E}[c_t^e - l_t]\right\}}, \tag{A35}$$

and

$$k^e = -(1 - \rho_w^e + \rho_c^e)\log(1 - \rho_w^e + \rho_c^e) - \rho_w^e\log(\rho_w^e) + \rho_c^e\log(\rho_c^e). \tag{A36}$$

Note that $\rho_w^e, \rho_c^e > 0$ because $W_t + L_t - C_t^e > 0$ along the optimal path.

Loglinear Euler equation

We can write equation (6.40) as

$$\begin{aligned} 1 &= \pi^e \; \mathrm{E}_t\left[\exp\left\{\log\delta - \gamma\left(c^e_{t+1} - c^e_t\right) + r_{i,t+1}\right\}\right] \\ &\quad + (1-\pi^e)\,\mathrm{E}_t\left[\exp\left\{\log\delta^r - \gamma\left(c^r_{t+1} - c^e_t\right) + r_{i,t+1}\right\}\right] \\ &\equiv \pi^e \; \mathrm{E}_t\left[\exp\left\{x_{t+1}\right\}\right] + (1-\pi^e)\; \mathrm{E}_t\left[\exp\left\{y_{t+1}\right\}\right] \;, \qquad \text{(A37)} \end{aligned}$$

where the notational correspondence between the first and second line is obvious, and $\delta^r \equiv (1-\pi^d)\delta$. Taking a second order Taylor expansion of $\exp\{x_{t+1}\}$ and $\exp\{y_{t+1}\}$ around $\overline{x}_t = \mathrm{E}_t[x_{t+1}]$ and $\overline{y}_t = \mathrm{E}_t[y_{t+1}]$ we can write:

$$\begin{aligned} 1 &\approx \pi^e \; \mathrm{E}_t\left[\exp\left\{\overline{x}_t\right\}\left(1 + (x_{t+1} - \overline{x}_t) + \frac{1}{2}(x_{t+1} - \overline{x}_t)^2\right)\right] \\ &\quad + (1-\pi^e)\; \mathrm{E}_t\left[\exp\left\{\overline{y}_t\right\}\left(1 + (y_{t+1} - \overline{y}_t) + \frac{1}{2}(y_{t+1} - \overline{y}_t)^2\right)\right] \\ &= \pi^e \exp\left\{\overline{x}_t\right\}\left(1 + \frac{1}{2}\;\mathrm{Var}_t\left(x_{t+1}\right)\right) + (1-\pi^e)\exp\left\{\overline{y}_t\right\}\left(1 + \frac{1}{2}\;\mathrm{Var}_t\left(y_{t+1}\right)\right) \;. \end{aligned}$$

Finally, a first-order Taylor expansion around zero gives

$$1 \approx \pi^e\left(1 + \overline{x}_t + \frac{1}{2}\;\mathrm{Var}_t\left(x_{t+1}\right)\right) + (1-\pi^e)\left(1 + \overline{y}_t + \frac{1}{2}\;\mathrm{Var}_t\left(y_{t+1}\right)\right),$$

or

$$\begin{aligned} 0 &= \pi^e\left(\log\delta - \gamma\mathrm{E}_t\left[c^e_{t+1} - c^e_t\right] + \mathrm{E}_t\left[r_{i,t+1}\right]\right. \qquad \text{(A38)} \\ &\quad \left. + \frac{1}{2}\mathrm{Var}_t\left[r_{i,t+1} - \gamma\left(c^e_{t+1} - c^e_t\right)\right]\right) \\ &\quad + (1-\pi^e)\left(\log\delta^r - \gamma\mathrm{E}_t\left[c^r_{t+1} - c^e_t\right] + \mathrm{E}_t\left[r_{i,t+1}\right]\right. \\ &\quad \left. \frac{1}{2}\mathrm{Var}_t\left[r_{i,t+1} - \gamma\left(c^r_{t+1} - c^e_t\right)\right]\right) . \end{aligned}$$

Optimal portfolio rule

We start guessing the functional form of the optimal policies in the employment state is:

$$\begin{aligned} c^e_t - l_t &= b^e_0 + b^e_1\left(w_t - l_t\right), \qquad \text{(A39)} \\ \alpha^e_t &= \alpha^e. \end{aligned}$$

Note that we can also write the optimal consumption policy in the retirement state (6.42) in the same form as equation (**??**):

$$c^r_{t+1} - l_{t+1} = b^r_0 + b^r_1 \left(w_{t+1} - l_{t+1} \right), \tag{A40}$$

where $b^r_1 = 1$.

Subtracting the log Euler equation (A38) for $r_{i.t+1} = r_f$ from the log Euler equation (A38) for $r_{i.t+1} = r_{t+1}$ yields:

$$\begin{aligned} \mathrm{E}_t\, r_{t+1} - r_f + \frac{1}{2} \mathrm{Var}_t \left(r_{t+1} \right) &= \gamma \pi^e \mathrm{Cov}_t \left(c^e_{t+1} - c^e_t, r_{t+1} \right) \\ &\quad + \gamma \left(1 - \pi^e \right) \mathrm{Cov}_t \left(c^r_{t+1} - c^e_t, r_{t+1} \right) \end{aligned} \tag{A41}$$

But equations (A39) and (A40), the log-linear intertemporal budget constraint (A33) and the trivial equality

$$c^s_{t+1} - c^e_t = (c^s_{t+1} - l_{t+1}) - (c^e_t - l_t) + (l_{t+1} - l_t), \tag{A42}$$

imply that

$$\begin{aligned} \mathrm{Cov}_t \left(c^s_{t+1} - c^e_t, r_{t+1} \right) &= \mathrm{Cov}_t \left(b^s_1 r^e_{p,t+1} + \left(1 - b^s_1 \right) \left(l_{t+1} - l_t \right), r_{t+1} \right) \\ &= b^s_1 \alpha^e_t \sigma^2_u + \left(1 - b^s_1 \right) \sigma_{\xi u}, \end{aligned} \tag{A43}$$

for $s = e$, r. The second line follows from the assumptions on asset returns and labor income. Substituting back into equation (A41) and using equation (A26) we find

$$\mu + \frac{1}{2} \sigma^2_u = \gamma \left[\left(\pi^e b^e_1 + \left(1 - \pi^e \right) \right) \alpha^e_t \sigma^2_u + \pi^e \left(1 - b^e_1 \right) \sigma_{\xi u} \right], \tag{A44}$$

from which equation (6.45) in text obtains immediately.

Optimal consumption rule

The log-Euler equation (A38) for $i = p$ and the trivial equality (A42) imply

$$\pi^e \mathrm{E}_t \left[c^e_{t+1} - l_{t+1} \right] + \left(1 - \pi^e \right) \mathrm{E}_t \left[c^r_{t+1} - l_{t+1} \right] = \left(c^e_t - l_t \right) + \Upsilon^e_t, \tag{A45}$$

where

$$\Upsilon^e_t \equiv \Upsilon^e = \frac{1}{\gamma} \left(\mathrm{E} \left[r^e_{p,t+1} \right] + \frac{1}{2} \mathrm{V}^e + \pi^e \log \delta + \left(1 - \pi^e \right) \log \delta^r \right) - g, \tag{A46}$$

and

$$\begin{aligned}\mathrm{V}^e &= [\pi^e(1-b_1^e\gamma)^2+(1-\pi^e)(1-b_1^r\gamma)^2]\,\mathrm{Var}[r_{p,t+1}^e]\\ &\quad +\pi^e\gamma(1-b_1^e)\,\mathrm{Var}[\Delta l_{t+1}]\\ &\quad -2\pi^e\gamma(1-\gamma b_1^e)(1-b_1^e)\,\mathrm{Cov}[r_{p,t+1}^e,\Delta l_{t+1}]. \end{aligned} \tag{A47}$$

If we substitute equations (A39) and (A40) into equation (A45) we obtain

$$\overline{b}_0+\overline{b}_1\,\mathrm{E}_t\left[w_{t+1}-l_{t+1}\right]=\Upsilon^e+b_0^e+b_1^e\left(w_t-l_t\right), \tag{A48}$$

where $\overline{b}_0=\pi^e b_0^e+(1-\pi^e)b_0^r$ and $\overline{b}_1=\pi^e b_1^e+(1-\pi^e)b_1^r$. Further substitution of the log budget constraint in the employment state (A33) and guess (A39) in the left-hand side of equation (A48) yields

$$\begin{aligned}&\overline{b}_0+\overline{b}_1\left(\rho_w^e-\rho_c^e b_1^e\right)(w_t-l_t)+\overline{b}_1\left(k^e-\rho_c^e b_0^e-g+\mathrm{E}\,r_{p,t+1}^e\right)\\ =\;&\Upsilon^e+b_0^e+b_1^e\left(w_t^e-l_t\right).\end{aligned}$$

Identifying coefficients on both sides of this equation we get the following two-equation system:

$$\begin{aligned}\overline{b}_1\left(\rho_w^e-\rho_c^e b_1^e\right) &= b_1^e,\\ \overline{b}_0+\overline{b}_1\left(k^e-\rho_c^e b_0^e-g+\mathrm{E}\,r_{p,t+1}^e\right) &= \Upsilon^e+b_0^e.\end{aligned}$$

We can solve this system recursively, since the first equation depends only on b_1^e and the second on b_1^e and b_0^e.

Simple algebraic manipulation of the first equation gives the following quadratic equation for b_1^e:

$$0=\pi^e\rho_c^e\left(b_1^e\right)^2+\left[1-\pi^e\rho_w^e+\left(1-\pi^e\right)\rho_c^e\right]b_1^e-\left(1-\pi^e\right)\rho_w^e. \tag{A49}$$

The expression for b_0^e is given by

$$\begin{aligned}b_0^e &= -\frac{1}{k_1}\left[\left(\frac{1}{\gamma}-\overline{b}_1\right)\mathrm{E}\left[r_{p,t+1}^e\right]+\frac{1}{\gamma}\left(\pi^e\log\delta+\left(1-\pi^e\right)\log\delta^r\right)\right.\\ &\quad\left.+\frac{1}{2\gamma}\,\mathrm{V}^e-\pi^e\left(1-b_1^e\right)g-\left(1-\pi^e\right)b_0^r-k^e\right],\end{aligned} \tag{A50}$$

with

$$k_1=\left(1-\pi^e\right)+\rho_c^e\overline{b}_1>0. \tag{A51}$$