CHAPTER 5

Risk Aversion and Investment Decisions, Part 1

Chapter Outline

5.1 Introduction

Chapters 3 and 4 provided a systematic procedure for assessing an investor's relative preference for various investment payoffs: rank them according to expected utility using a VNM-utility representation constructed to reflect the investor's preferences over random payments. The subsequent postulate of risk aversion further refined this idea: it is natural to hypothesize that the utility-of-money function entering the investor's VNM index is concave ($U''(\,) < 0$). Two widely used measures were introduced and interpreted, each permitting us to assess an investor's degree of risk aversion. In the setting of a zero-cost investment paying either $(+h)$ or $(-h)$, these measures were shown to be linked with the

Intermediate Financial Theory.

minimum probability of success above one-half necessary for a risk-averse investor to take on such a prospect willingly. They differ only as to whether (h) measures an absolute amount of money or a proportion of the investors' initial wealth.

In this chapter, we begin to use these ideas with a view toward understanding the two most important financial decisions an investor will make. The first is his portfolio composition decision: what are his relative demands for assets of different risk classes and, in particular, his demand for risk-free versus risky assets? This topic is dealt with in Sections 5.2–5.5. The second concerns an investor's consumption/savings decision: how much of an investor's current income should he allocate to savings and how much to consumption? When an investor saves, he augments his wealth in the future which implies extending our modeling context to multiple periods. This is the focus of Section 5.6.

A multiperiod setting naturally gives rise to preference phenomena not present in the atemporal context of earlier chapters. In Section 5.7, we review a number of departures from the strict VNM-expected utility paradigm that allows these phenomena to be given formal representation.

5.2 Risk Aversion and Portfolio Allocation: Risk-Free Versus Risky Assets

5.2.1 The Canonical Portfolio Problem

Consider an investor with wealth level Y_0, who is deciding what amount, a, to invest in a risky portfolio with uncertain rate of return $\tilde{r}$. We can think of the risky asset as being, in fact, the market portfolio under the "old" capital asset pricing model (CAPM), to be reviewed in Chapter 8. The alternative is to invest in a risk-free asset that pays a certain rate of return r_f. The time horizon is one period. The investor's wealth at the end of the period is given by

$$\tilde{Y}_1 = (1 + r_f)(Y_0 - a) + a(1 + \tilde{r}) = Y_0(1 + r_f) + a(\tilde{r} - r_f)$$

The choice problem he must solve can be expressed as

$$\max_a EU(\tilde{Y}_1) = \max EU(Y_0(1 + r_f) + a(\tilde{r} - r_f)) \tag{5.1}$$

where $U(\ )$ is his utility-of-money function and E the expectations operator.

This formulation of the investor's problem is fully in accord with the lessons of the prior chapter. Each choice of a leads to a different uncertain payoff distribution, and we want to find the choice that corresponds to the most preferred such distribution. By construction of his VNM representation, this is the payoff pattern that maximizes his expected utility.

Under risk aversion ($U''(\ ) < 0$), the necessary and sufficient first-order condition (FOC) for problem (5.1) is given by

$$E[U'(Y_0(1+r_f)+a(\tilde{r}-r_f))(\tilde{r}-r_f)]=0 \tag{5.2}$$

Analyzing Eq. (5.2) allows us to describe the relationship between the investor's degree of risk aversion and his portfolio's composition as per the following theorem.

Theorem 5.1 Assume $U'(\ ) > 0$, and $U''(\ ) < 0$ and let $\hat{a}$ denote the solution to problem (5.1). Then

$$\hat{a} > 0 \Leftrightarrow E\tilde{r} > r_f$$
$$\hat{a} = 0 \Leftrightarrow E\tilde{r} = r_f$$
$$\hat{a} < 0 \Leftrightarrow E\tilde{r} < r_f$$

Proof Since this is a fundamental result, it is worthwhile to make clear its (straightforward) justification. We follow the argument presented in Arrow (1971), Chapter 2.

Define $W(a) = E\{U(Y_0(1+r_f)+a(\tilde{r}-r_f))\}$. The FOC (5.2) can then be written $W'(a) = E[U'(Y_0(1+r_f)+a(\tilde{r}-r_f))(\tilde{r}-r_f)] = 0$. By risk aversion ($U'' < 0$), $W''(a) = E[U''(Y_0(1+r_f)+a(\tilde{r}-r_f))(\tilde{r}-r_f)^2] < 0$, i.e., $W'(a)$ is everywhere decreasing. It follows that $\hat{a}$ will be positive if and only if $W'(0) = U'(Y_0(1+r_f))E(\tilde{r}-r_f) > 0$ (since then a will have to be increased from the value of 0 to achieve equality in the FOC). Since U' is always strictly positive, this implies $\hat{a} > 0$ if and only if $E(\tilde{r}-r_f) > 0$.

The other assertions follow similarly.

Theorem 5.1 asserts that a risk-averse agent will invest in the risky asset or portfolio only if the expected return on the risky asset exceeds the risk-free rate. From another perspective, a risk-averse agent will *always* participate (possibly via an arbitrarily small stake) in a risky investment when the odds are favorable.

5.2.2 Illustration and Examples

It is worth pursuing the above result to get a sense of how large a is relative to Y_0. Our findings will, of course, be preference dependent. Let us begin with the fairly standard and highly tractable utility function $U(Y) = \ln Y$. For added simplicity, let us also assume that the risky asset is forecast to pay either of two returns (corresponding to an "up" or "down" stock market), $r_2 > r_1$, with probabilities π and $1-\pi$, respectively. It makes sense (why?) to assume $r_2 > r_f > r_1$ and $E\tilde{r} = \pi r_2 + (1-\pi)r_1 > r_f$.

Under this specification, the FOC (5.2) becomes

$$E\left\{\frac{\tilde{r}-r_{\mathrm{f}}}{Y_0(1+r_{\mathrm{f}})+a(\tilde{r}-r_{\mathrm{f}})}\right\}=0$$

Writing out the expectation explicitly yields

$$\frac{\pi(r_2-r_{\mathrm{f}})}{Y_0(1+r_{\mathrm{f}})+a(r_2-r_{\mathrm{f}})}+\frac{(1-\pi)(r_1-r_{\mathrm{f}})}{Y_0(1+r_{\mathrm{f}})+a(r_1-r_{\mathrm{f}})}=0$$

which, after some straightforward algebraic manipulation, gives

$$\frac{a}{Y_0}=\frac{-(1+r_{\mathrm{f}})[E\tilde{r}-r_{\mathrm{f}}]}{(r_1-r_{\mathrm{f}})(r_2-r_{\mathrm{f}})}>0 \tag{5.3}$$

This is an intuitive sort of expression: the fraction of wealth invested in risky assets increases with the return premium paid by the risky asset ($E\tilde{r}-r_{\mathrm{f}}$) and decreases with an increase in the return dispersion around r_{f} as measured by (r_2-r_{f}) $(r_{\mathrm{f}}-r_1)$.[1]

Suppose $r_{\mathrm{f}}=0.05$, $r_2=0.40$, and $r_1=-0.20$, and $\pi=\frac{1}{2}$ (the latter information guarantees $E\tilde{r}=0.10$). In this case, $a/Y_0=0.6$: 60% of the investor's wealth turns out to be invested in the risky asset. Alternatively, suppose $r_2=0.30$ and $r_1=-0.10$ (same r_{f}, π, and $E\tilde{r}$); here we find that $a/Y_0=1.4$. This latter result must be interpreted to mean that an investor would prefer to invest at least his full wealth in the risky portfolio. If possible, he would even want to borrow an additional amount, equal to 40% of his initial wealth, at the risk-free rate, and invest this amount in the risky portfolio as well. In comparing these two examples, we see that the return dispersion is much smaller in the second case (lower risk in a mean−variance sense) with an unchanged return premium. With less risk and unchanged mean returns, it is not surprising that the proportion invested in the risky asset increases very substantially. We will see, however, that, somewhat surprisingly, this result does not generalize without further assumption on the form of the investor's preferences.

5.3 Portfolio Composition, Risk Aversion, and Wealth

In this section, we consider how an investor's portfolio decision is affected by his degree of risk aversion and his wealth level. A natural first exercise is to compare the portfolio composition across individuals of differing risk aversion. The answer to this first question conforms with intuition: if John is more risk averse than Amos, he optimally invests a smaller fraction of his wealth in the risky asset. This is the essence of our next two theorems.

[1] That this fraction is independent of the wealth level is not a general result, as we shall find out in Section 5.3.

Theorem 5.2 (**Arrow, 1971**) Suppose, for all wealth levels Y, $R_A^1(Y) > R_A^2(Y)$, where $R_A^i(Y)$ is the measure of absolute risk aversion of investor $i, i = 1, 2$. Then $\hat{a}_1(Y) < \hat{a}_2(Y)$.

That is, the more risk-averse agent, as measured by his absolute risk-aversion measure, will always invest less in the risky asset, given the same level of wealth. This result does not depend on measuring risk aversion via the absolute Arrow–Pratt measure. Indeed, since $R_A^1(Y) > R_A^2(Y) \Leftrightarrow R_R^1(Y) > R_R^2(Y)$, Theorem 5.2 can be restated as

Theorem 5.3 Suppose, for all wealth levels $Y > 0$, $R_R^1(Y) > R_R^2(Y)$ where $R_R^i(Y)$ is the measure of relative risk aversion of investor i, $i = 1, 2$. Then $\hat{a}_1(Y) < \hat{a}_2(Y)$.

Continuing with the example of Section 5.2.2, suppose now that the investor's utility function has the form $U(Y) = (Y^{1-\gamma})/(1-\gamma)$, $\gamma > 1$. This utility function displays both greater absolute and greater relative risk aversion than $U(Y) = \ln Y$ (you are invited to prove this statement). From Theorems 5.2 and 5.3, we would expect this greater risk aversion to manifest itself in a reduced willingness to invest in the risky portfolio. Let us see if this is the case.

For these preferences, the expression corresponding to Eq. (5.3) is

$$\frac{a}{Y_0} = \frac{(1+r_f)\left\{[(1-\pi)(r_f - r_1)]^{\frac{1}{\gamma}} - (\pi(r_2 - r_f))^{\frac{1}{\gamma}}\right\}}{(r_1 - r_f)\{\pi(r_2 - r_f)\}^{\frac{1}{\gamma}} - (r_2 - r_f)\{(1-\pi)(r_f - r_1)\}^{\frac{1}{\gamma}}} \tag{5.4}$$

In the case of our first example, but with $\gamma = 3$, we obtain, by simple direct substitution,

$$\frac{a}{Y_0} = 0.24$$

indeed, only 24% of the investor's assets are invested in the risky portfolio, down from 60% earlier.

The next logical question is to ask how the investment in the risky asset varies with the investor's total wealth as a function of his degree of risk aversion. Let us begin with statements appropriate to the absolute measure of risk aversion.

Theorem 5.4 (**Arrow, 1971**) Let $\hat{a} = \hat{a}(Y_0)$ be the solution to Problem (5.1), then:

i. $R_A'(Y) < 0 \Leftrightarrow \hat{a}'(Y_0) > 0$
ii. $R_A'(Y) = 0 \Leftrightarrow \hat{a}'(Y_0) = 0$
iii. $R_A'(Y) > 0 \Leftrightarrow \hat{a}'(Y_0) < 0$

Case (i) is referred to as declining absolute risk aversion (DARA). Agents with this property become more willing to accept greater bets as they become wealthier. Theorem 5.4 says that such agents will also increase the amount invested in the risky asset ($\hat{a}'(Y_0 > 0)$. To state matters slightly differently, an agent with the indicated DARA will, if he becomes wealthier, be willing to put some of that additional wealth at risk. Utility functions of this form are quite common: those considered in the example, $U(Y) = \ln Y$ and $U(Y) = (Y^{1-\gamma})/(1-\gamma)$, $\gamma > 0$, display this property. It also makes intuitive sense.

Under constant absolute risk aversion or CARA, case (ii), the amount invested in the risky asset is unaffected by the agent's wealth. This result is somewhat counterintuitive. One might have expected that a CARA decision maker, in particular one with little risk aversion, would invest some of his or her increase in initial wealth in the risky asset. Theorem 5.4 disproves this intuition. An example of a CARA utility function is $U(Y) = -e^{-\nu Y}$. Indeed,

$$R_A(Y) = \frac{-U''(Y)}{U'(Y)} = \frac{-(-\nu^2)e^{-\nu Y}}{\nu e^{-\nu Y}} = \nu$$

Let's verify the claim of Theorem 5.4 for this utility function. Consider

$$\max_a E\left(-e^{-\nu(Y_0(1+r_f)+a(\tilde{r}-r_f))}\right)$$

The FOC is

$$E\left[\nu(\tilde{r}-r_f)e^{-\nu(Y_0(1+r_f)+a(\tilde{r}-r_f))}\right] = 0$$

Now compute da/dY_0; by differentiating the above equation, we obtain

$$E\left[\nu(\tilde{r}-r_f)e^{-\nu(Y_0(1+r_f)+a(\tilde{r}-r_f))}\left(1 + r_f + (\tilde{r}-r_f)\frac{da}{dY_0}\right)\right] = 0$$

$$(1+r_f)\underbrace{E\left[\nu(\tilde{r}-r_f)e^{-\nu(Y_0(1+r_f)+a(\tilde{r}-r_f))}\right]}_{=0 \text{ (by the FOC)}} + E\left[\underbrace{\nu(\tilde{r}-r_f)^2}_{>0}\frac{da}{dY_0}\underbrace{e^{-\nu(Y_0(1+r_f)+a(\tilde{r}-r_f))}}_{>0}\right] = 0$$

therefore, $da/dY_0 \equiv 0$.

For the above preference ordering, and our original two-state risky distribution,

$$\hat{a} = \frac{1}{\nu}\left(\frac{1}{r_1 - r_2}\right)\ln\left(\frac{(1-\pi)}{\pi}\left(\frac{r_f - r_1}{r_2 - r_f}\right)\right)$$

Note that in order for $\hat{a}$ to be positive, it must be that

$$0 < \frac{(1-\pi)}{\pi}\left(\frac{r_f - r_1}{r_2 - r_f}\right) < 1$$

A sufficient condition is that $\pi > 1/2$.

Case (iii) is one with increasing absolute risk aversion (IARA). It says that as an agent becomes wealthier, he reduces his investments in risky assets. This does not make much sense, and we will generally ignore this possibility. Note, however, that the quadratic utility function, which is of some significance as we will see later on, possesses this property.

Let us now think in terms of the relative risk-aversion measure. Since it is defined for bets expressed as *a proportion of wealth*, it is appropriate to think in terms of *elasticities*, or of how the fraction invested in the risky asset changes as wealth changes. Define $\eta(Y, \hat{a}) = (\mathrm{d}\hat{a}/\hat{a})/(dY/Y) = (Y/\hat{a})(\mathrm{d}\hat{a}/dY)$, i.e., the wealth elasticity of investment in the risky asset. For example, if $\eta(Y, \hat{a}) > 1$, as wealth Y increases, the **percentage** increase in the amount optimally invested in the risky portfolio exceeds the percentage increase in Y. Or as wealth increases, the **proportion** optimally invested in the risky asset increases. Analogous to Theorem 5.4 is Theorem 5.5.

Theorem 5.5 (**Arrow, 1971**) If, for all wealth levels Y,

i. $R_R'(Y) = 0$ (CRRA) then $\eta = 1$
ii. $R_R'(Y) < 0$ (DRRA) then $\eta > 1$
iii. $R_R'(Y) > 0$ (IRRA) then $\eta < 1$

In his article, Arrow gives support for the hypothesis that the rate of relative risk aversion should be constant with constant relative risk aversion (CRRA) ≈ 1. In particular, it can be shown that if an investor's utility of wealth is to be bounded above as wealth tends to ∞, then $\lim_{Y \to \infty} R'_R(Y) \geq 1$; similarly, if $U(Y)$ is to be bounded below as $Y \to 0$, then $\lim_{Y \to 0} R'_R(Y) \leq 1$. These results suggest that if we wish to assume CRRA, then CRRA = 1 is the appropriate value.[2] Utility functions of the CRRA class include $U(Y) = (Y^{1-\gamma})/(1-\gamma)$, where $R_R(Y) = \gamma$ and $R_A(Y) = \gamma/Y$.

5.4 Special Case of Risk-Neutral Investors

As noted in Chapter 4, a risk-neutral investor is one who does not care about risk; he ranks investments solely on the basis of their expected returns. The utility function of such an

[2] Note that the above comments also suggest the appropriateness of weakly increasing relative risk aversion as an alternative working assumption.

agent is necessarily of the form $U(Y) = c + \mathrm{d}Y, c, d$ constants, $d > 0$. (Check that $U'' = 0$ in this case.)

What proportion of his wealth will such an agent invest in the risky asset? The answer is: provided $E\tilde{r} > r_f$ (as we have assumed), **all** of his wealth will be invested in the risky asset.[3] This is clearly seen from the following. Consider the agent's portfolio problem:

$$\max_a E(c + d(Y_0(1 + r_f) + a(\tilde{r} - r_f)) = \max_a [c + d(Y_0(1 + r_f)) + da(E\tilde{r} - r_f)]$$

With $E\tilde{r} > r_f$ and, consequently, $d(E\tilde{r} - r_f) > 0$, this expression is increasing in a. This means that if the risk-neutral investor is unconstrained, he will attempt to borrow as much as possible at r_f and reinvest the proceeds in the risky portfolio. He is willing, without bound, to exchange certain payments for uncertain claims of greater expected value. As such he stands willing to absorb all of the economy's financial risk. If we specify that the investor is prevented from borrowing, then the maximum will occur at $a = Y_0$.

5.5 Risk Aversion and Risky Portfolio Composition

So far we have considered the question of how an investor should allocate his wealth between a risk-free asset and a risky asset or portfolio. We now go one step further and ask the following question: when is the *composition* of the portfolio (i.e., the percentage of the portfolio's value invested in each of the J risky assets that compose it) independent of the agent's wealth level? This question is particularly relevant in light of current investment practices whereby portfolio decisions are usually taken in steps. Step 1, often associated with the label "asset allocation," is the choice of instruments: stocks, bonds, and riskless assets (possibly alternative investments as well, such as hedge funds, private equity, and real estate); Step 2 is the country or sector allocation decision: here the issue is to optimize not across asset class but across geographical regions or industrial sectors. Step 3 consists of the individual stock-picking decisions made on the basis of information provided by financial analysts. The issuing of asset and country/sector allocation "grids" by all major financial institutions, tailored to the risk profile of the different clients, but independent of their wealth levels (and of changes in their wealth), is predicated on the hypothesis that differences in wealth (across clients) and changes in their wealth do not require adjustments in portfolio composition provided risk tolerance is either unchanged or controlled for.

Let us illustrate the issue in more concrete terms; take the example of an investor with invested wealth equal to \$12,000 and optimal portfolio proportions of $a_1 = \frac{1}{2}$, $a_2 = \frac{1}{3}$, and $a_3 = \frac{1}{6}$ (only three assets are considered). In other words, this individual' portfolio holdings are \$6000 in asset 1, \$4000 in asset 2, and \$2000 in asset 3. The implicit assumption

[3] One way to interpret Theorem 5.1 in this light is to infer that it suggests investors are approximately risk neutral over small gambles (Arrow, 1971). Yet this is apparently not the case (c.f. (110, −100; ½)).

behind the most common asset management practice is that, were the investor's wealth to double to \$24,000, the new optimal portfolio would naturally be

$$\text{Asset 1: } \frac{1}{2}(\$24,000) = \$12,000$$

$$\text{Asset 2: } \frac{1}{3}(\$24,000) = \$8,000$$

$$\text{Asset 3: } \frac{1}{6}(\$24,000) = \$4,000$$

The question we pose in the present section is: Is this hypothesis supported by theory? The answer is generally no, in the sense that it is only for very specific preferences (utility functions) that the asset allocation is optimally left unchanged in the face of changes in wealth levels. Fortunately, these specific preferences include some of the major utility representations. The principal result in this regard is as follows.

Theorem 5.6 (Cass and Stiglitz, 1970) Let the vector $\begin{pmatrix} \hat{a}_1(Y_0) \\ \vdots \\ \hat{a}_J(Y_0) \end{pmatrix}$ denote the amount optimally invested in the J risky assets if the wealth level is Y_0.

Then $\begin{pmatrix} \hat{a}_1(Y_0) \\ \vdots \\ \hat{a}_J(Y_0) \end{pmatrix} = \begin{pmatrix} a_1 \\ \vdots \\ a_J \end{pmatrix} f(Y_0)$

(for some positive function $f(\cdot)$) if and only if either

i. $U'(Y_0) = (\theta\, Y_0 + \kappa)^{\Delta}$ or

ii. $U'(Y_0) = \xi\, e^{-\nu Y_0}$

There are, of course, implicit restrictions on the choice of θ, κ, Δ, ξ, and υ to ensure, in particular, that $U''(Y_0) < 0$.[4]

Integrating (i) and (ii), respectively, in order to recover the utility functions corresponding to these marginal utilities, one finds, significantly, that the first includes the CRRA class of functions:

$U(Y_0) = \frac{1}{1-\gamma} Y_0^{(1-\gamma)}$ $\gamma \neq 1$, and $U(Y_0) = \ln(Y_0)$, while the second corresponds to the CARA:

$$U(Y_0) = \frac{\xi}{-\nu} e^{-\nu Y_0}$$

[4] For (i), we must have **either** $\theta > 0$, $\Delta < 0$, and Y_0 such that $\theta Y_0 + \kappa \geq 0$ **or** $\theta < 0$, $\kappa < 0$, $\Delta > 0$, and $Y_0 \leq -\frac{\kappa}{\theta}$. For (ii), $\xi > 0$, $-\upsilon < 0$, and $Y_0 \geq 0$.

In essence, Theorem 5.6 states that it is only in the case of utility functions satisfying constant absolute or constant relative risk-aversion preferences (and some generalization of these functions of minor interest) that the relative composition of the risky portion of an investor's optimal portfolio is invariant to changes in his wealth.[5] Only in these cases should the investor's portfolio composition be left unchanged as invested wealth increases or decreases. It is only with such utility specifications that the standard "grid" approach to portfolio investing is formally justified.[6]

5.6 Risk Aversion and Savings Behavior

5.6.1 Savings and the Riskiness of Returns

We have thus far considered the relationship between an agent's degree of risk aversion and the composition of his portfolio. This was accomplished in an atemporal one-period context. A related, though significantly different, question is to ask how an agent's **savings rate** is affected by an increase in the degree of risk facing him. Note that the savings question will be the first occasion where we must explicitly introduce a time dimension into the analysis (i.e., where the important trade-off involves the present versus the future). It is to be expected that the answer to this question will be influenced, in a substantial way, by the agent's degree of risk aversion.

Consider first an agent solving the following two-period consumption−savings problem:

$$\begin{aligned} &\max_{s} E\{U(Y_0 - s) + \delta U(s\tilde{R})\} \\ &\text{s.t.} \quad Y_0 \geq s \geq 0 \end{aligned} \tag{5.5}$$

where Y_0 is initial (period zero) wealth, s is the amount saved and entirely invested in a risky portfolio with uncertain gross risky return, $\tilde{R} = 1 + \tilde{r}$, $U(\ )$ is the agent's period utility-of-consumption function, and δ is his subjective discount factor.[7] The subjective discount rate $\delta < 1$ captures the extent to which the investor values consumption in the future less than current consumption.

5 As noted earlier, the constant absolute risk-aversion class of preferences has the property that the total amount invested in risky assets is invariant to the level of wealth. It is not surprising therefore that the proportionate allocation among the available risky assets is similarly invariant as this theorem asserts.

6 Theorem 5.6 does not mean, however, that the fraction of initial wealth invested in the risk-free asset versus the risky "mutual fund" is invariant to changes in Y_0. The CARA class of preferences discussed in the previous footnote is a case in point.

7 Note that this $U(c) = U(Y_0 - s)$ is in principle no different from the (indirect) utility of *wealth* function considered earlier in Chapter 3: the income remaining after savings is the income available for consumption.

The FOC for this problem (assuming an interior solution) is given by

$$U'(Y_0 - s) = \delta E(U'(s\tilde{R})\tilde{R}\} \tag{5.6}$$

It is clear from the above equation that the properties of the return distribution $\tilde{R}$ will influence the optimal level of s. One is particularly interested to know how optimal savings is affected by the riskiness of returns.

To be concrete, let us think of two return distributions $\tilde{R}_A$, $\tilde{R}_B$ such that $\tilde{R}_B$ is riskier than $\tilde{R}_A$ and $E\tilde{R}_A = E\tilde{R}_B$. From our previous work (Theorem 4.4), we know this can be made precise by stating that $\tilde{R}_A$ SSD $\tilde{R}_B$ or that $\tilde{R}_B$ is a mean-preserving spread of $\tilde{R}_A$. In other words, one can write $\tilde{R}_B = \tilde{R}_A + \tilde{\varepsilon}$, where $\tilde{\varepsilon}$ is a zero mean random variable uncorrelated with $\tilde{R}_A$. Let s_A and s_B be, respectively, the savings out of Y_0 corresponding to the return distributions $\tilde{R}_A$ and $\tilde{R}_B$. The issue is whether s_A is larger than s_B or if the converse is true. In other words, can one predict that a representative risk-averse agent *will save more or less* when confronted with riskier returns on his or her savings?

Let us try to think intuitively about this issue. On the one hand, one may expect that more risk will mean a decrease in savings because "a bird in the hand is worth two in the bush." This can be viewed as a substitution effect: A riskier return can be likened to an increase in the cost of future consumption. A rational individual may then opt to consume more today. On the other hand, a risk-averse individual may want to increase savings in the face of uncertainty as a precautionary measure, in order to guarantee a minimum standard of living in the future. This reaction indeed is associated with the notion of "precautionary savings." The reader is invited to verify that this ambiguity is resolved in a mean−variance world in favor of the first argument. In that context, riskier returns imply a decrease in the RHS of Eq. (5.6), or a decrease in the expected future marginal utility of consumption weighted by the gross return. For the equality to be restored, consumption today must increase, and, consequently, savings must decrease.

It is important to realize, however, that the mean−variance response is not representative of the reactions of all risk-averse agents. Indeed, observers seeking to explain the increase in the US personal savings rate post 2007 have regularly pointed to the rising uncertainties surrounding the macroeconomic situation in general and the pace of economic growth in particular.[8] As our discussion suggests, the key technical issue is whether the RHS of Eq. (5.6) is increased or decreased by an increase in risk. Applying reasoning similar to that used when discussing risk aversion (see Section 5.2), it is easy to see that this issue, in fact, revolves around whether the RHS of Eq. (5.6) (i.e., $\delta U'(sR)R \equiv sg(R)$), is convex (in which case it increases) or concave (in which case it decreases) in R.

[8] Which, if they are right, would tend to suggest that "the world is not mean−variance."

Suppose, to take an extreme case, that the latter is linear in R; we know that linearity means that the RHS of Eq. (5.6) can be written as $\delta E(g(R)) = \delta g(ER)$. But since R_A and R_B have the same mean, this implies that the RHS of Eq. (5.6), and consequently optimal savings, are unaffected by an increase in risk. If, on the other hand, $g(R)$ is concave, then $E(g(R)) < g(E(R))$; this is the previous case with a concave $g(\cdot)$, upward deviations from the mean produce smaller changes in the values attained by the function than downward deviations. It is then the case that the RHS will be decreased as a result of the mean-preserving spread on R. The reverse is true if $g(\ )$ is convex.

Note that in the all important case where $U(c) = \ln(c)$, $g(R)$ is in fact a constant function of R, with the obvious result that the savings decision is not affected by the increase in the riskiness of the return distribution. This difference in results between two of the workhorses of finance (mean variance and log utility) is worth emphasizing.

Let us now compute the second derivative of $g(R)$:[9]

$$g''(R) = 2U''(sR)s + s^2RU'''(sR) \tag{5.7}$$

Using Eq. 5.7, one can express the sign of g'' in terms of the relative rate of risk aversion as in Theorem 5.7.

Theorem 5.7 (**Rothschild and Stiglitz, 1971**) Let $\tilde{R}_A$ and $\tilde{R}_B$ be two return distributions with identical means such that $\tilde{R}_A$ SSD $\tilde{R}_B$, and let s_A and s_B be, respectively, the savings out of Y_0 corresponding to the return distributions $\tilde{R}_A$ and $\tilde{R}_B$.

If $R_R'(Y) \leq 0$ and $R_R(Y) > 1$, then $s_A < s_B$
If $R_R'(Y) \geq 0$ and $R_R(Y) < 1$, then $s_A > s_B$

Proof To prove this assertion, we need the following Lemma 5.7.

Lemma 5.7 $R_R'(Y)$ has the same sign as $-[U'''(Y)Y + U''(Y)(1 + R_R(Y))]$.

Proof Since $R_R(Y) = \dfrac{-YU''(Y)}{U'(Y)}$

$$R_R'(Y) = \frac{[-U'''(Y)Y - U''(Y)]U'(Y) - [-U''(Y)Y]U''(Y)}{[U'(Y)]^2}$$

[9] $g(R) = U'(sR)R \Rightarrow g'(R) = U''(sR)sR + U'(sR)$ and $g''(R)$ as in Eq. (5.6). In the log case, we get $g(R) = 1/s$ and thus $g'(R) = g''(R) = 0$.

Since $U'(Y) > 0$, $R'_R(Y)$ has the same sign as

$$\frac{[-U'''(Y)Y - U''(Y)]U'(Y) - [-U''(Y)Y]U''(Y)}{U'(Y)}$$
$$= -U'''(Y)Y - U''(Y) - \left[\frac{-U''(Y)Y}{U'(Y)}\right]U''(Y)$$
$$= -\{U'''(Y)Y + U''(Y)[1 + R_R(Y)]\}$$

Now we can proceed with the theorem. We will show only the first implication as the second follows similarly. By the lemma, since $R'_R(Y) < 0$,

$$-\{U'''(Y)Y + U''(Y)[1 + R_R(Y)]\} < 0, \quad \text{or}$$
$$\{U'''(Y)Y + U''(Y)[1 + R_R(Y)]\} > 0$$

In addition, since $U''(Y) < 0$ and $R_R(Y) > 1$,

$$U'''(Y)Y + U''(Y)(2) > \{U'''(Y)Y + U''(Y)[1 + R_R(Y)]\} > 0$$

This is true for all Y; hence

$2U''(sR) + sRU'''(sR) > 0$. Multiplying left and right by $s > 0$, one gets $2U''(sR)s + s^2RU'''(sR) > 0$, which by Eq. (5.6) implies

$$g''(R) > 0$$

But by the earlier remarks, this means that $s_A < s_B$ as required.

Theorem 5.7 implies that for the class of constant relative risk-aversion utility functions, i.e., functions of the form

$$U(c) = (1-\gamma)^{-1}c^{1-\gamma}$$

($0 < \gamma \neq 1$), an increase in risk increases savings if $\gamma > 1$ and decreases it if $\gamma < 1$, with the $U(c) = \ln(c)$ case being the watershed for which savings is unaffected. For broader classes of utility functions, this theorem provides a partial characterization only, suggesting different investors react differently according to whether they display declining or increasing relative risk aversion.

A more complete characterization of the issue of interest is afforded if we introduce the concept of prudence, first proposed by Kimball (1990). Let

$\mathbf{P}(c) = (-U'''(c))/(U''(c))$ be a measure of absolute prudence, while by analogy with risk aversion,
$c\mathbf{P}(c) = (-cU'''(c))/(U''(c))$ then measures relative prudence. Theorem 5.7 can now be restated as Theorem 5.8.

Theorem 5.8 Let $\tilde{R}_A$ and $\tilde{R}_B$ be two return distributions such that $\tilde{R}_A$ SSD $\tilde{R}_B$, and let s_A and s_B be, respectively, the savings out of Y_0 corresponding to the return distributions $\tilde{R}_A$ and $\tilde{R}_B$. Then,

$$s_A > s_B \text{ iff } c\mathbf{P}(c) \leq 2$$

and conversely,

$$s_A < s_B \text{ iff } c\mathbf{P}(c) > 2$$

That is, risk-averse individuals with relative prudence lower than 2 decrease savings while those with relative prudence above 2 increase savings in the face of an increase in the riskiness of returns.

Proof We have seen that $s_A < s_B$ if and only if $g''(R) > 0$. From Eq. (5.6), this means

$$sRU'''(sR)/U''(sR) < s - 2, \text{ or}$$

$$c\mathbf{P}(c) = \frac{sRU'''(sR)}{-U''(sR)} > 2$$

as claimed. The other part of the proposition is proved similarly.

5.6.2 Illustrating Prudence

The relevance of the concept of prudence can be illustrated in the simplest way if we turn to a slightly different problem, where one ignores uncertainty in returns (assuming, in effect, that the net return is identically zero) while asking how savings in period zero is affected by uncertain labor income in period 1. Our remarks in this context are drawn from Kimball (1990).

Let us write the agent's second period labor income, Y, as $Y = \overline{Y} + \tilde{Y}$, where $\overline{Y}$ is the mean labor income and $\overline{Y}$ measures deviations from the mean (of course, $E\tilde{Y} = 0$). The simplest form of the decision problem facing the agent is thus:

$$\max_s E\{U(Y_0 - s) + \beta U(s + \overline{Y} + \tilde{Y})]$$

where $s = s_i$ satisfies the FOC.

i. $U'(Y_0 - s_i) = \beta E(U'(s_i + \overline{Y} + \tilde{Y})\}$
   It will be of interest to compare the solution s_i to the above FOC with the solution to the analogous decision problem, denoted s_{ii}, in which the uncertain labor income component is absent. The latter FOC is simply

ii. $U'(Y_0 - s_{ii}) = \beta U'(s_{ii} + \overline{Y})$

The issue once again is whether and to what extent s_i differs from s_{ii}.

One approach to this question, which gives content to the concept of prudence, is to ask what the agent would need to be paid (what compensation is required in terms of period 2 income) to ignore labor income risk—in other words, for his first-period consumption and savings decision to be unaffected by uncertainty in labor income. The answer to this question leads to the definition of the *compensating precautionary premium* $\psi = \psi(\overline{Y}, \tilde{Y}, s)$ as the amount of additional second period wealth (consumption) that must be given to the agent in order that the solution to (i) coincides with the solution to (ii). That is, the compensatory precautionary premium $\psi(\overline{Y}, \tilde{Y}, s)$ is defined as the solution of

$$U'(Y_0 - s_{ii}) = \beta E\{U'(s_{ii} + \overline{Y} + \tilde{Y} + \psi(\overline{Y}, \tilde{Y}, s))\}$$

Kimball (1990) proves the following two results.

Theorem 5.9 Let $U(\ )$ be three times continuously differentiable and $\mathbf{P}(s)$ be the index of absolute prudence. Then

i. $\psi(\overline{Y}, \tilde{Y}, s) \approx 1/2\sigma^2_{\tilde{Y}}\ \mathbf{P}(s + \overline{Y})$

Furthermore, let $U_1(\ )$ and $U_2(\ )$ be two second-period utility functions for which

$$\mathbf{P}_1(s) = \frac{-U'''_1(s)}{U''_1(s)} < \frac{-U'''_2(s)}{U''_2(s)} = \mathbf{P}_2(s), \text{ for all } s$$

Then

ii. $\psi_2(\overline{Y}, \tilde{Y}, s) > \psi_1(\overline{Y}, \tilde{Y}, s)$ for all $s, \overline{Y}, \tilde{Y}$

Theorem 5.9 (i) shows that investors' precautionary premia are directly proportional to the product of their prudence index and the variance of their uncertain income component, a result analogous to the characterization of the measure of absolute risk aversion obtained in Section 4.3. The result of Theorem 5.9 (ii) confirms the intuition that the more "prudent" the agent, the greater the compensating premium.

5.6.3 The Joint Saving–Portfolio Problem

Although for conceptual reasons we have so far distinguished the consumption–savings and the portfolio allocation decisions, it is obvious that the two decisions must be considered jointly. We now formalize the consumption/savings/portfolio allocation problem:

$$\max_{\{a,s\}} U(Y_0 - s) + \delta EU(s(1 + r_f) + a(\tilde{r} - r_f)) \tag{5.8}$$

where s denotes the total amount saved and a is the amount invested in the risky asset. Specializing the utility function to the form $U(Y) = (Y^{1-\gamma})/(1-\gamma)$, the FOCs for this joint decision problem are

$$s : (Y_0 - s)^{-\gamma}(-1) + \delta E([s(1+r_f) + a(\tilde{r} - r_f)]^{-\gamma}(1+r_f)) = 0$$
$$a : E[(s(1+r_f) + a(\tilde{r} - r_f))^{-\gamma}(\tilde{r} - r_f)] = 0$$

The first equation spells out the condition to be satisfied at the margin for the savings level—and by corollary, consumption—to be optimal. It involves comparing the marginal utility today with the expected marginal utility tomorrow, with the rate of transformation between consumption today and consumption tomorrow being the product of the discount factor and the gross risk-free return. This FOC need not occupy us any longer here. The interesting element is the solution to the second FOC: it has the exact same form as Eq. (5.2) with the endogenous (optimal) s replacing the exogenous initial wealth level Y_0. Let us rewrite this equation as

$$s^{-\gamma} E\left[\left((1+r_f) + \frac{a}{s}(\tilde{r} - r_f)\right)^{-\gamma}(\tilde{r} - r_f)\right] = 0$$

which implies

$$E\left[\left((1+r_f) + \frac{a}{s}(\tilde{r} - r_f)\right)^{-\gamma}(\tilde{r} - r_f)\right] = 0$$

This equation confirms the lessons of Eqs. (5.3) and (5.4): For the selected utility function, the proportion of savings invested in the risky asset is independent of s, the amount saved. This is an important result: while it does not generalize to other utility functions, it nevertheless opens up the possibility of a straightforward extension of the savings–portfolio problem to many periods. We pursue this important extension in Chapter 16.

5.7 Generalizing the VNM-Expected Utility Representation

Project evaluation exercises and the household's consumption–savings problem are fundamentally intertemporal in nature. In an intertemporal context, however, various preference phenomena provide additional challenges to the VNM-expected utility framework. It is these extensions that we propose to address in the remaining section of this chapter.

While proving to be extremely useful in explaining real-world phenomena, the extensions we consider are not "radical." In particular, preferences continue to be defined over money payoffs rather than gains or losses relative to some benchmark, and all proposed extensions reduce to our standard VNM-expected utility set up for specific choices of the relevant parameters.

5.7.1 Preferences for the Timing of Uncertainty Resolution

Under the VNM-expected utility representation, investors are assumed to be concerned only with actual final payoffs and the cumulative probabilities of attaining them. In particular, they are assumed to be indifferent to the timing of uncertainty resolution. To get a better idea of what this means in a choice-theoretic context, consider the two investment payoff trees depicted in Figure 5.1. These investments are to be evaluated from the viewpoint of date 0 (today).

Under the expected utility postulates, these two payoff structures would be valued (in utility terms) identically as

$$EU(\tilde{P}) = U(100) + \delta[\pi U(150) + (1 - \pi)U(25)]$$

where δ is the investor's subjective discount factor.

This means that a VNM investor would not care if the uncertainty were resolved in period 1 or one period later.[10] Yet, people are, in fact, very different in this regard. Some want to know the outcome of an uncertain event as soon as possible; others prefer to postpone it as long as possible.[11]

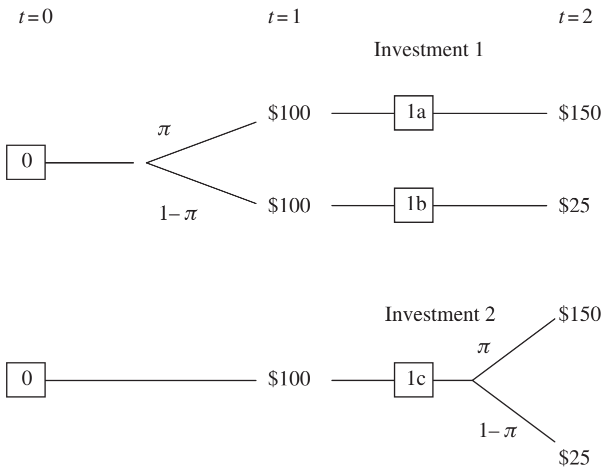

Figure 5.1
Two investments: payoff patterns for earlier versus later resolution of uncertainty.

[10] In particular, preference for the timing of uncertainty resolution violates VNM-expected utility axiom C.1a (see Chapter 3).

[11] One commonplace illustration of these differences arises in persons' differing attitudes "to seeing the doctor." When confronted with an unforeseen pain, some individuals immediately visit their physician to have it precisely diagnosed. Others let the days pass in the hope the pain will "go away" of its own accord. Only if it persists will they (reluctantly) see their physician for a diagnosis. There are, of course, many other reasons for avoiding the doctor, such as the inability to pay his fee.

Kreps and Porteus (1978) were the first to develop a theory that allowed for preferences for the timing of uncertainty resolution. They showed that if investor preferences over uncertain sequential payoffs were of the form

$$U_0(P_1, P_2(\tilde{\theta})) = W(P_1, E(U_1(P_1, P_2(\tilde{\theta})))$$

then investors would prefer early (late) resolution of uncertainty according to whether $W(P_1,.)$ is convex (concave) (loosely, whether $W_{22} > 0$ or $W_{22} < 0$). In the above representation P_i is the payoff in period $i = 1,2$. If $W(P_1, \,)$ were concave, for example, the expected utility of investment 1 would be lower than investment 2.

Let us illustrate this distinction using the choices in Figure 5.1. We assume functional forms similar to those used in an illustration of Kreps and Porteus (1978); in particular, assume $W(P_1, EU) = EU^{1.5}$, and $U_1(P_1, P_2(\tilde{\theta})) = (P_1 + P_2(\tilde{\theta}))^{1/2}$. Let $\pi = 0.5$ and note that the overall composite function $U_0(\,)$ is concave in all of its arguments. In computing utilities at the decision nodes [0], [1a], [1b], and [1c] (the latter decisions are trivial ones), we must be especially scrupulous to observe exactly the dates at which the uncertainty is resolved under the two alternatives:

[1a]: $EU_1^{1a}(P_1, P_2(\theta)) = (100 + 150)^{1/2} = 15.811$

[1b]: $EU_1^{1b}(P_1, P_2(\theta)) = (100 + 25)^{1/2} = 11.18$

[1c]: $EU_1^{1c}(P_1, P_2(\theta)) = 0.5(100 + 150)^{1/2} + 0.5(100 + 25)^{1/2} = 13.4955$

At $t = 0$, the expected utility on the upper tree is

$$\begin{aligned} EU_0^{1a,1b}(P_1, P_2(\tilde{\theta})) &= EW^{1a,1b}(P_1, P_2(\tilde{\theta})) \\ &= 0.5W(100, 15.811) + 0.5W(100, 11.18) \\ &= 0.5(15.811)^{1.5} + 0.5(11.18)^{1.5} \\ &= 50.13 \end{aligned}$$

while on the lower tree

$$EU_0^{1c}(P_1, P_2(\tilde{\theta})) = W(100, 13.4955) = (13.4955)^{1.5} = 49.57$$

This investor clearly prefers early resolution of uncertainty, which is consistent with the convexity of the $W(\,)$ function. Note that the result is simply an application of Jensen's inequality.[12] If $W(\,)$ had been concave, the ordering would be reversed.

Recent empirical evidence (e.g., Brown and Kim (2013)) suggests that a majority of individuals prefer early resolution of uncertainty, a fact that may have special significance

[12] Let $a = (100 + 150)^{1/2}$, $b = (100 + 25)^{1/2}$, $g(x) = x^{1.5}$ (convex), $EU_0{}^{1a,1b}\,(P, \tilde{P}_2(\theta)) = Eg(x) > g(Ex) = EU_0{}^{1c}(P_1, \tilde{P}_2(\theta))$, where $x = a$ with prob $= 0.5$ and $x = b$ with prob $= 0.5$.

for the pricing of stocks.[13] The intuition goes along the following lines: Bansal and Yaron (2004) demonstrate that the aggregate US consumption and dividend (for the CRSP index of US stocks) data series are consistent with both having a common, small, highly persistent mean growth component.[14] This means (e.g., in the case of dividends) there will be long periods of high average dividend growth followed by long periods where average dividend growth is low, and vice versa. Because of the high persistence, these series reflect the late resolution of uncertainty. When combined with preferences for early resolution of uncertainty on the part of market participants, this persistence phenomenon leads to low equity prices in an equilibrium asset pricing context (see Chapter 10): investors will only hold late-resolution-of-uncertainty equity securities if their prices are very low. In fact, in the Bansal and Yaron (2004) equilibrium asset pricing context average equity returns are high enough (equity prices are low enough) to resolve the equity premium puzzle (see Chapter 2).[15]

The functional forms used in the prior illustrative example are not widely used. In Section 5.7.3, however, we introduce the popular Epstein-Zin (1991) utility specification which, depending on the choice of parameter values, can also display preference for the early or late uncertainty resolution.

5.7.2 Preferences That Guarantee Time-Consistent Planning

The notion of *time-consistent planning* is this: if, at each date, the agent could plan against any future contingency, what is the required relationship among the family of orderings $\{\succeq_t : t = 0, 1, 2, \ldots, T\}$ that will cause plans that were optimal with respect to preferences $\succeq_0$ to remain optimal in all future time periods given all that may happen in the interim (i.e., intermediate consumption experiences and the specific way uncertainty has evolved)? In particular, what utility function representation will guarantee this property?

When considering decision problems over time, such as portfolio investments over a multiperiod horizon, time consistency would appear to be a desirable property. In its absence, one would observe portfolio rebalancing not motivated by any outside event or information flow, but simply resulting from the inconsistency of the date t preference ordering of the investor compared with the preferences on which her earlier portfolio positions were based. Asset trades would then be fully motivated by endogenous and unobservable preference issues and would thus be basically unexplainable.

13 For the experimental results presented in Brown and Kim (2013), "a majority (60.4%) of subjects demonstrate a preference for early resolution (of uncertainty), about one-third are indifferent (36.6%), and 3% exhibit a preference for late resolution" (Brown and Kim, 2013, p. 2).

14 The CRSP index comprises nearly all traded USA stocks and is compiled by the Center for Research in Security Prices at the University of Chicago.

15 We explore this topic in greater detail in Chapter 10.

To see what it takes for a utility function to be time consistent, let us consider two periods where at date 1 any one of $\theta \in S$ possible states of nature may be realized. Let c_0 denote a possible consumption level at date 0, and let $c_1(\theta)$ denote a possible consumption level in period 1 if state "θ" occurs. Johnsen and Donaldson (1985) demonstrate that if initial preferences $\succeq_0$, with utility representation $U(\ )$, are to guarantee time-consistent planning, there must exist continuous and monotone increasing functions $f(\ )$ and $\{U_\theta(.,.): \theta \in S\}$ such that

$$U(c_0, c_1(\theta){:}\theta \in S) = f(c_0, U_\theta(c_0, c_1(\theta)){:}\theta \in S) \tag{5.9}$$

where $U_\theta(.,.)$ is the state θ contingent utility function.[16]

This result means the utility function must be of a form such that the utility representations in future states can be recursively nested as individual arguments of the overall utility function. This condition is satisfied by the VNM-expected utility form

$$U(c_0, c_1(\theta){:}\theta \in S) = U_0(c_0) + \sum_\theta \pi_s U(c_1(\theta)) \tag{5.10}$$

which clearly display the structure of Eq. (5.9).[17] The VNM-utility representation is thus time consistent, but the latter property can also be accommodated by more general utility functions which are not VNM. To see this, consider the following specialization of representation (5.9), where there are three possible states at $t = 1$:

$$U(c_0, c_1(1), c_1(2), c_1(3)) = \left\{c_0 + \pi_1 U_1(c_0, c_1(1)) + [\pi_2 U_2(c_0, c_1(2))]^{\frac{1}{3}} \pi_3 U_3(c_0, c_1(3))\right\}^{\frac{1}{2}} \tag{5.11}$$

where

$$U_1(c_0, c_1(1)) = \log(c_0 + c_1(1)),$$

$$U_2(c_0, c_1(2)) = (c_0)^{\frac{1}{2}}(c_1(2))^{\frac{1}{2}}, \text{ and}$$

$$U_3(c_0, c_1(3)) = c_0 c_1(3)$$

[16] With the utility-of-money function defined here over state-contingent consumption, we are implicitly assuming one composite consumption good with a normalized price of one in every period. State-contingent consumption, $c_t(\theta)$, thus also represents state contingent income $Y_t(\theta)$ available for spending after savings have been set aside. The consumption−income identity legitimizes our use of the utility-of-money function $U(\ )$ as simultaneously representing a utility of consumption function.

[17] The many period extension (slightly generalized) of Eq. (5.10) is to postulate that investor preferences over contingent consumption plans assume the form $E_0\left(\sum_{t=0}^{T} \delta^t u(c_t)\right)$, $0 < \delta < 1$.
At any future date $t > 0$, the operative ordering thus becomes $E_t\left(\sum_{j=0}^{T-t} \delta^{t+j} u(c_{t+j})\right) = \delta^t E_0\left(\sum_{j=0}^{T-t} \delta^j u(c_{t+j})\right)$ which, by the ordinal property of utility functions, is equivalent to Eq. (5.9). Accordingly, preferences of this form always admit time consistent planning.

In this example, preferences are clearly not linear in the probabilities and thus they are not of the VNM-expected utility type. Nevertheless, identification (5.11) is of the form of Eq. (5.9). It also has the feature that preferences in any future state are independent of irrelevant alternatives, where the irrelevant alternatives are those consumption plans for states that do not occur. As such, agents with these preferences will never experience regret, and the Allais paradox will not be operational.

Consistency of choice makes sense and turns out to be important for individual and national savings behavior, but is it borne out empirically? Unfortunately, the answer is: frequently not. A simple illustration of this is a typical pure-time-preference experiment from the psychology literature (uncertainty in future states is not even needed). Participants are asked to choose among the following monetary prizes:[18]

Question 1: Would you prefer \$100 today or \$200 in 2 years?
Question 2: Would you prefer \$100 in 6 years or \$200 in 8 years?

Respondents often prefer the \$100 in Question 1 and the \$200 in Question 2, not realizing that Question 2 involves the same choice as Question 1 but with a 6-year delay. If these people behave true to their answers, they will be time inconsistent: in the case of Question 2, although they state their preference now for the \$200 prize in 8 years, when year 6 arrives they will take the \$100 and run! Is there a way this choice reversal can be conveniently represented in a conventional, utility-based framework? Providing an answer to this question constitutes our next topic.

5.7.2.1 Quasi-Hyperbolic Discounting

The commonplace reversal cited above suggests very different attitudes toward *intended* saving in the future versus actual savings today. This is apparent from a slight reframing of the two questions posed above into a consumption–savings context.

Question 1′: Would you prefer \$100 today or the opportunity to save and invest the \$100 for an assured payoff in 2 years of \$200?
Question 2′: Would you prefer \$100 in 6 years or the opportunity to save and invest the \$100 for 2 more years for an assured payoff of \$200?

While Questions 1′ and 2′ are framed slightly differently from Questions 1 and 2, a study of savings patterns would reveal similar responses: \$100 today preferred to \$200 in 2 years, yet investing the \$100 in 6 years being preferred to receiving the \$100 at that time. In other words, investors apparently are willing to commit to savings in the future while being less willing to save today: there is a bias for "immediate gratification." A convenient way to

---

[18] See Ainslie and Haslan (1992) for details. Similar illustrations may be found in Thaler (1981).

summarize these choices formally within the general expected utility framework is to define a family of preference orderings $\{\mathbb{U}_t(\ ){:}t = 0, 1, 2, \ldots\}$ by

$$\mathbb{U}_t(\cdots) = E_t\left\{u(c_t) + \beta \sum_{j=1}^{T} \delta^{t+j} u(c_{t+j})\right\} \tag{5.12}$$

where $\beta < 1$.[18] Note that for such an investor, his consumption/savings decision between periods $t + j$ and $t + j + 1$, $j > 0$, as planned today (period t) is governed by the subjective discount factor δ. Yet when period t transpires, his consumption savings behavior will be governed by the subjective discount factor $\delta\beta < \delta$. It is as though the investor has a dual personality. Individuals of this persuasion will behave dynamically time inconsistently: they will plan to save a lot in the future, whereas "when the future arrives" they end up saving very much less. Following Laibson (1997) preferences of the form (5.12) are now referred to as quasi-hyperbolic.

In situations where preferences are of the form (5.12), the ability of the individual consumer−investor to achieve an appropriate life-cycle consumption/savings plan becomes an issue of the need for self-control in the form of a supporting strategy of "commitment." By a commitment strategy, we mean one by which "self t" binds his "future selves" to save or retain assets in excess of what they would otherwise voluntarily accomplish when they themselves become the decision makers. In practice, commitment investing frequently takes the form of investing in illiquid assets, assets for which the early sale is either forbidden, except in very narrowly defined circumstances of ill-health, or limited by costly penalties for early withdrawal. Savings in the form of contributions to employer-sponsored retirement accounts, IRAs, and even short-term Christmas-club accounts (recently reintroduced by Walmart) are commonplace illustrations. The acquisition of a home financed by a mortgage contract is the most widespread of commitment devices: not only is a home difficult to sell (liquidate) quickly but its associated mortgage payments become a form of required savings. Laibson (1997) illustrates the notion of commitment in the context of illiquid versus liquid asset acquisition in a formal game-theoretic (current self versus future selves) consumption−savings context.

Phelps and Pollak (1968) originally used preferences of the form (5.12) to explain historically low US savings rates in the United States. Using analysis based on preference ordering (5.12) Laibson (1997) similarly argues that the decline in the US savings rate beginning in the 1980s may have been due to the proliferation of unsecured debt in the

[18] This family of ordering was first proposed by Phelps and Pollak (1968) and later resurrected and further developed by Laibson (1997).

form of easily accessible credit cards.[19] Already low national savings rates can be lowered even further when financial liberalization allows the commitment devices noted above to be circumvented. The increased ease of obtaining home equity loans (which allow the partial liquidation of real estate wealth) in the years leading up to the financial crisis is a case in point. We are also reminded of the work of Jappelli and Pagano (1994) (Chapter 1), who demonstrated that financial deregulation in Italy led to a reduction in the Italian savings rate. Although the focus of this deregulation was the removal of financial constraints, in some respects commitment devices were thereby removed as well.

Sections 5.7.1 and 5.7.2 have focused largely on time preference related issues. In the final segment of our discussion, we bring back risk preferences and, in particular, consider the joint interaction of time and risk preferences.

5.7.3 Separating Risk and Time Preferences

Consider the standard consumption−savings problem (5.5), and suppose once again that the agent's period utility function has been specialized to have the standard CRRA form,

$$U(c) = \frac{Y^{1-\gamma}}{1-\gamma}, \quad \gamma > 0$$

For this utility function, the single-parameter γ captures not only the agent's sensitivity to atemporal risk, but also his sensitivity to consumption variation across time periods or, equivalently, his willingness to substitute consumption in one period for consumption in another. A high γ signals a strong desire to avoid atemporal consumption risk and, simultaneously, a strong reluctance to substitute consumption in one period for consumption in another. To see this more clearly, consider a deterministic version of Problem (5.5) where $\delta < 1$, $\tilde{R} \equiv 1$:

$$\max_{0 \le s \le Y_0} \{U(Y_0 - s) + \delta U(s)\}$$

The necessary and sufficient FOC

$$-(Y_0 - s)^{-\gamma} + \delta s^{-\gamma} = 0 \text{ or}$$

$$\left(\tfrac{1}{\delta}\right)^{\frac{1}{\gamma}} = \left(\frac{Y_0 - s}{s}\right)$$

[19] The national savings rate is important because it largely determines an economy's long-run capital stock available to each of its workers, and thus per worker productivity and consumption. Economies that save more will, ceteris paribus, also enjoy lower interest rates. See Web Chapter A.

With $\delta < 1$, as the agent becomes more and more risk averse ($\gamma \mapsto \infty$), $((Y_0 - s)/s) = (c_0/c_1) \mapsto 1$; i.e., $c_0 \approx c_1$. For this preference structure, a highly risk-averse agent will also seek an intertemporal consumption profile that is very smooth.

We have stressed repeatedly the pervasiveness of the preference for smooth consumption whether across time or across states of nature, and its relationship with the notion of risk aversion. It is time to recognize that while in an atemporal setting a desire for smooth consumption across states of nature is the very definition of risk aversion, in a multiperiod environment, risk aversion and the desire for intertemporal consumption smoothing should not necessarily be equated. After all, one may speak of intertemporal consumption smoothing in a no-risk, deterministic setting, and one may speak of risk aversion in an uncertain, atemporal environment. The situation considered so far where the same parameter determines both is thus restrictive. Indeed, empirical studies tend to suggest that typical individuals are more averse to intertemporal substitution (they desire very smooth consumption intertemporally) than they are averse to risk *per se*. This latter fact cannot be captured in the aforementioned, single-parameter setting.

Is it possible to generalize the standard utility specification and break this coincidence of time and risk preferences? Epstein and Zin (1989, 1991) answer positively and propose a class of utility functions that allows each dimension to be parameterized separately while still preserving the time consistency property discussed in Section 5.7.2. They provide, in particular, the axiomatic basis for preferences over lotteries leading to the Kreps and Porteus (1978)-like utility representation:

$$U_t = U(c_t, c_{t+1}, c_{t+2}, \ldots) = W(c_t, \mathrm{CE}(\tilde{U}_{t+1})) \tag{5.13}$$

where

i. $U_t = U(c_t, \tilde{c}_{t+1}, \tilde{c}_{t+2}, \ldots)$ describes the investor's period t future lifetime utility associated with consumption stream $(c_t, \tilde{c}_{t+1}, \tilde{c}_{t+2}, \ldots)$
ii. $W(\,,\,)$ is an aggregator function analogous to the $f(\;)$ function in Eq. (5.9); and
iii. $\mathrm{CE}(\tilde{U}_{t+1})$ is the period t certainty equivalent of uncertain lifetime utility beginning at date $t+1$, measured in terms of period $t+1$ consumption units.

Epstein and Zin (1989) and Weil (1989) explore the following CES (constant elasticity of intertemporal substitution)-like specialized version of Eq. (5.13):

$$W(c_t, \mathrm{CE}(\tilde{U}_{t+1})) = \left[(1-\delta)c_t^{1-\rho} + \delta\left(\mathrm{CE}(\tilde{U}_{t+1})\right)^{1-\rho}\right]^{\frac{1}{1-\rho}}, \quad \rho \neq 1 \text{ or} \tag{5.14a}$$

$$W(c_t, \mathrm{CE}(\tilde{U}_{t+1})) = (1-\delta)\log c_t + \delta \log(\mathrm{CE}(\tilde{U}_{t+1})), \quad \rho = 1 \tag{5.14b}$$

where $1/\rho$ is the elasticity of intertemporal substitution. Since a higher ρ means the investor prefers a smoother intertemporal, certain consumption stream, ρ is viewed as the investor's time preference parameter.

The certainty equivalent of future lifetime utility, $\mathrm{CE}(\tilde{U}_{t+1})$, is computed in a manner directly analogous to what was done in Sections 4.4 and 4.5. It thus identifies γ as the risk preference parameter, where γ and ρ may be chosen independently of one another.

$$\left[\mathrm{CE}(\tilde{U}_{t+1})\right] = \left(E_t\left(\tilde{U}_{t+1}^{1-\gamma}\right)\right)^{\frac{1}{1-\gamma}}, 1 \neq \gamma > 0, \text{ or} \tag{5.15a}$$

$$\log \mathrm{CE}(\tilde{U}_{t+1}) = E_t(\log \tilde{U}_{t+1}), \quad \gamma = 1 \tag{5.15b}$$

By Eq. (5.13), $U_t(\ )$ is recursively determined and, in general, has no time-separable representation. Accordingly, Epstein and Zin (1989) preferences are not easy to use. In a consumption savings context, for example, the lifetime utility function U_t, the savings decision s_t and the portfolio allocation decision (risk free versus risky assets) must all be determined simultaneously and endogenously.[20]

As for preference for the timing of uncertainty resolution, if $\gamma > \rho$, it can be demonstrated that the investor prefers early resolution and vice versa if $\gamma < \rho$. If $\gamma = \rho$, recursive substitution of U_t yields

$$U_t = \left[(1-\delta)E_t \sum_{j=0}^{\infty} \delta^j c_{t+j}^{1-\gamma}\right]^{\frac{1}{1-\gamma}}$$

which represents the same preference as

$$E_t \sum_{j=0}^{\infty} \delta^j c_{t+j}^{1-\gamma}$$

and is thus equivalent to the usual time-separable expected utility case with CRRA utility. In general Eqs. (5.11)–(5.12) are not of the expected utility form as probabilities do not enter linearly.

As amply demonstrated in Weil (1989), Epstein-Zin (1989) preferences, in and of themselves, will not resolve the equity premium puzzle. As noted at the close of Section 5.7.1, further structure must be placed on the manner in which consumption uncertainty is modeled (see Chapter 10).

5.8 Conclusions

We have considered, in a very simple context, the relationship between an investor's degree of risk aversion, on the one hand, and his desire to save and the composition of his portfolio

[20] See van Binsbergen et al. (2008) for one numerical solution technique; also see the web notes to this chapter.

on the other. Most of the results were intuitively acceptable, and that, in itself, makes us more confident of the VNM representation.

Are there any lessons here for portfolio managers to learn? At least three lessons are suggested:

1. Regardless of the level of risk, some investment in risky assets is warranted, even for the most risk-averse clients (provided $E\tilde{r} > r_f$). This is the substance of Theorem 5.1.
2. As the value of a portfolio changes significantly, the asset allocation (proportion of wealth invested in each asset class) and the risky portfolio composition should be reconsidered. How that should be done depends critically on the client's attitudes toward risk. This is the substance of Theorems 5.4–5.6.
3. Investors are willing, in general, to pay to reduce income (consumption) risk and would like to enter into mutually advantages transactions with institutions less risk averse than themselves. The extreme case of this is illustrated in Section 5.5.

We went on to consider how greater return uncertainty influences savings behavior. On this score and in some other instances, this chapter has illustrated the fact that, somewhat surprisingly, risk aversion is not always a sufficient hypothesis to recover intuitive behavior in the face of risk. The third derivative of the utility function often plays a role. The notion of prudence permits an elegant characterization in these situations.

We concluded the chapter by considering three plausible modifications of the general expected utility framework in order to capture phenomena not representable under the strict VNM-expected utility axioms. As we will see in future chapters, preference representations allowing for both the separation of time and risk parameters and a preference for the early timing of uncertainty resolution, will be fundamental to a resolution of the equity premium phenomenon.

In many ways, this chapter has aimed at providing a broad perspective allowing us to place modern portfolio theory and its underlying assumptions in their proper context. We are now prepared to revisit this pillar of modern finance.

References

Ainslie, G., Haslan, N., 1992. Hyperbolic discounting. In: Lowenstein, G., Elster, J. (Eds.), Choice Over Time. Russell Sage Foundation, New York, NY.

Arrow, K.J., 1971. Essays in the Theory of Risk Bearing. Markham, Chicago, IL.

Bansal, R., Yanon, A., 2004. Risks for the long run: a potential resolution of asset pricing puzzles. J. Finan. 59, 1481–1509.

Brown, A., Kim, H., 2013, Do individuals have preferences used in macro-finance models? An experimental investigation, Working Paper, Texas A&M University, Department of Economics.

Cass, D., Stiglitz, J.E., 1970. The structure of investor preference and asset returns and separability in portfolio allocation: a contribution to the pure theory of mutual funds. J. Econ. Theory. 2, 122–160.

Epstein, L.G., Zin, S.E., 1989. Substitution, risk aversion, and the temporal behavior of consumption growth and asset returns I: theoretical framework. Econometrica. 57, 937−969.

Epstein, L.G., Zin, S.E., 1991. Substitution, risk aversion, and the temporal behavior of consumption growth and asset returns II: an empirical analysis. J. Empir. Econ. 99, 263−286.

Jappelli, T., Pagano, M., 1994. Savings, growth, and liquidity constraints. Q. J. Econ. 109, 83−109.

Johnsen, T., Donaldson, J.B., 1985. The structure of intertemporal preferences under uncertainty and time consistent plans. Econometrica. 53, 1451−1458.

Kimball, M.S., 1990. Precautionary savings in the small and in the large. Econometrica. 58, 53−73.

Kreps, D., Porteus, E., 1978. Temporal resolution of uncertainty and dynamic choice theory. Econometrica. 46, 185−200.

Laibson, D., 1997. Golden eggs and hyperbolic discounting. Q. J. Econ. 112, 443−477.

Phelps, E.S., Pollak, R.A., 1968. On second best national savings and game-equilibrium growth. Rev. Econ. Stud. 35, 185−199.

Rothschild, M., Stiglitz, J., 1971. Increasing risk II: its economic consequences. J. Econ. Theory. 3, 66−85.

Thaler, R., 1981. Some empirical evidence on dynamic inconsistency. Econ. Lett. 8, 201−207.

van Binsbergen, J., Fernandez-Villaverde, J., Koijen, R., Rubio-Ramirez, J., 2008. Working with Epstein−Zin preferences: computation and likelihood estimation of DSGE models with recursive preferences, Working Paper, Duke University.

Weil, P.h., 1989. The equity premium puzzle and the risk free rate puzzle. J. Monet. Econ. 24, 401−421.