CHAPTER 5

Portfolio Choice Problems

Michael W. Brandt
Fuqua School of Business, Duke University, Durham, NC

Contents

Abstract

This chapter is devoted to the econometric treatment of portfolio choice problems. The goal is to describe, discuss, and illustrate through examples the different econometric approaches proposed in the literature for relating the theoretical formulation and solution of a portfolio choice problem to the data. In focusing on the econometrics of the portfolio choice problem, this chapter is at best a cursory overview of the broad portfolio choice literature. In particular, much of the discussion is focused on the single period portfolio choice problem with standard preferences, normally distributed returns, and frictionless markets. There are many recent advances in the portfolio choice literature, some cited below but many regrettably omitted, that relax one or more of these simplifying assumptions. The econometric techniques discussed in this chapter can be applied to these more realistic formulations. The chapter is divided into three parts. Section 2 reviews the theory of portfolio choice in discrete and continuous time. It also discusses a number of modeling issues and extensions that arise in formulating the problem. Section 3 presents the two traditional econometric approaches to portfolio choice problems: plug-in estimation and Bayesian decision theory. In Section 4, I then describe a more recently developed econometric approach for drawing inferences about optimal portfolio weights without modeling return distributions.

Keywords: portfolio choice; discrete time; continuous time; plug-in estimation; Bayesian decision theory; optimal portfolio weights

1. INTRODUCTION

After years of relative neglect in academic circles, portfolio choice problems are again at the forefront of financial research. The economic theory underlying an investor's optimal portfolio choice, pioneered by Markowitz (1952), Merton (1969, 1971), Samuelson (1969), and Fama (1970), is by now well understood. The renewed interest in portfolio choice problems follows the relatively recent empirical evidence of time-varying return distributions (e.g., predictability and conditional heteroskedasticity) and is fueled by realistic issues including model and parameter uncertainty, learning, background risks, and frictions. The general focus of the current academic research is to identify key aspects of real-world portfolio choice problems and to understand qualitatively as well as quantitatively their role in the optimal investment decisions of individuals and institutions.

Whether for academic researchers studying the portfolio choice implications of return predictability, for example, or for practitioners whose livelihood depends on the outcome of their investment decisions, a critical step in solving realistic portfolio choice problems is to relate the theoretical formulation of the problem and its solution to the data. There are a number of ways to accomplish this task, ranging from calibration with only vague regard for the data to decision theoretic approaches which explicitly incorporate the specification of the return model and the associated statistical inferences in the investor's decision process. Surprisingly, given the practical importance of portfolio choice problems, no single econometric approach has emerged yet as clear favorite. Because each approach has its advantages and disadvantages, an approach favored in one context is often less attractive in another.

This chapter is devoted to the econometric treatment of portfolio choice problems. The goal is to describe, discuss, and illustrate through examples the different econometric approaches proposed in the literature for relating the theoretical formulation and solution of a portfolio choice problem to the data. The chapter is intended for academic researchers who seek an introduction to the empirical implementation of portfolio choice problems as well as for practitioners as a review of the academic literature on the topic. In focusing on the econometrics of the portfolio choice problem, this chapter is at best a cursory overview of the broad portfolio choice literature. In particular, much of the discussion is focused on the single period portfolio choice problem with standard preferences, normally distributed returns, and frictionless markets. There are many recent advances in the portfolio choice literature, some cited below but many regrettably omitted, that relax one or more of these simplifying assumptions. The econometric techniques discussed in this chapter can be applied to these more realistic formulations.

The chapter is divided into three parts. Section 2 reviews the theory of portfolio choice in discrete and continuous time. It also discusses a number of modeling issues and extensions that arise in formulating the problem. Section 3 presents the two traditional econometric approaches to portfolio choice problems: plug-in estimation and Bayesian decision theory. In Section 4, I then describe a more recently developed econometric approach for drawing inferences about optimal portfolio weights without modeling return distributions.

2. THEORETICAL PROBLEM

2.1. Markowitz Paradigm

The mean–variance paradigm of Markowitz (1952) is by far the most common formulation of portfolio choice problems. Consider N risky assets with random return vector R_{t+1} and a riskfree asset with known return R_t^f. Define the *excess* returns $r_{t+1} = R_{t+1} - R_t^f$ and denote their conditional means (or risk premia) and covariance matrix by μ_t and Σ_t, respectively. Assume, for now, that the excess returns are i.i.d. with constant moments.

Suppose the investor can only allocate wealth to the N risky securities. In the absence of a risk-free asset, the mean–variance problem is to choose the vector of portfolio weights x, which represent the investor's relative allocations of wealth to each of the N risky assets, to minimize the variance of the resulting portfolio return $R_{p,t+1} = x' R_{t+1}$ for a predetermined target expected return of the portfolio $R_t^f + \overline{\mu}$:

$$\min_x \operatorname{var}[R_{p,t+1}] = x' \Sigma x, \tag{2.1}$$

subject to

$$\mathrm{E}[R_{p,t+1}] = x'(R^f + \mu) = (R^f + \overline{\mu}) \quad \text{and} \quad \sum_{i=1}^{N} x_i = 1. \tag{2.2}$$

The first constraint fixes the expected return of the portfolio to its target, and the second constraint ensures that all wealth is invested in the risky assets. Setting up the Lagrangian and solving the corresponding first-order conditions (FOCs), the optimal portfolio weights are as follows:

$$x^\star = \Lambda_1 + \Lambda_2 \overline{\mu} \tag{2.3}$$

with

$$\Lambda_1 = \frac{1}{D}\left[B(\Sigma^{-1}\iota) - A(\Sigma^{-1}\mu)\right] \quad \text{and} \quad \Lambda_2 = \frac{1}{D}\left[C(\Sigma^{-1}\mu) - A(\Sigma^{-1}\iota)\right], \tag{2.4}$$

where ι denotes an appropriately sized vector of ones and where $A = \iota' \Sigma^{-1} \mu$, $B = \mu' \Sigma^{-1} \mu$, $C = \iota' \Sigma^{-1} \iota$, and $D = BC - A^2$. The minimized portfolio variance is equal to $x^{\star\prime} \Sigma x^\star$.

The Markowitz paradigm yields two important economic insights. First, it illustrates the effect of diversification. Imperfectly correlated assets can be combined into portfolios

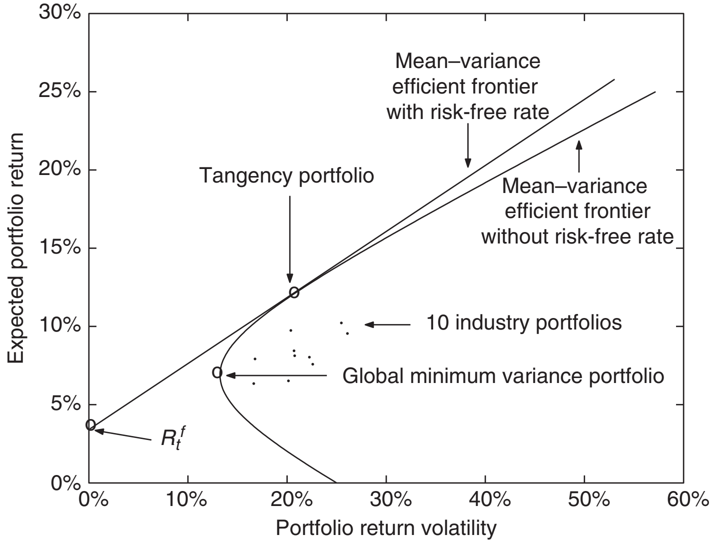

Figure 5.1 Mean–variance frontiers with and without risk-free asset generated by historical moments of monthly returns on 10 industry-sorted portfolios. Expected return and volatility are annualized.

with preferred expected return-risk characteristics. Second, the Markowitz paradigm shows that, once a portfolio is fully diversified, higher expected returns can only be achieved through more extreme allocations (notice $x^{\star}$ is linear in $\overline{\mu}$) and therefore by taking on more risk.

Figure 5.1 illustrates graphically these two economic insights. The figure plots as hyperbola the mean–variance frontier generated by the historical moments of monthly returns on 10 industry-sorted portfolios. Each point on the frontier gives along the horizonal axis the minimized portfolio return volatility (annualized) for a predetermined expected portfolio return (also annualized) along the vertical axis. The dots inside the hyperbola represent the 10 individual industry portfolios from which the frontier is constructed. The fact that these dots lie well inside the frontier illustrates the effect of diversification. The individual industry portfolios can be combined to generate returns with the same or lower volatility and the same or higher expected return. The figure also illustrates the fundamental trade-off between expected return and risk. Starting with the least volatile portfolio at the left tip of the hyperbola (the global minimum variance portfolio), higher expected returns can only be achieved at the cost of greater volatility.

If the investor can also allocate wealth to the risk-free asset, in the form of unlimited risk-free borrowing and lending at the risk-free rate R_t^f, any portfolio on the mean–variance frontier generated by the risky assets (the hyperbola) can be combined with the risk-free asset on the vertical axis to generate an expected return-risk profile that lies on a straight line from the risk-free rate (no risky investments) through the frontier portfolio (fully invested in risky asset) and beyond (leveraged risky investments). The optimal combination of the risky frontier portfolios with risk-free borrowing and lending is the one that maximizes the Sharpe ratio of the overall portfolio, defined as $\mathrm{E}[r_{p,t+1}]/\mathrm{std}[r_{p,t+1}]$ and represented graphically by the slope of the line from the risk-free asset through the risky frontier portfolio. The highest obtainable Sharpe ratio is achieved by the upper tangency on the hyperbola shown in Fig. 5.1. This tangency therefore represents the mean–variance frontier with risk-free borrowing and lending. The critical feature of this mean–variance frontier with risk-free borrowing and lending is that every investor combines the risk-free asset with the *same* portfolio of risky assets – the tangency portfolio in Fig. 5.1.

In the presence of a risk-free asset, the investor allocates fractions x of wealth to the risky assets and the remainder $(1 - \iota' x)$ to the risk-free asset. The portfolio return is therefore $R_{p,t+1} = x' R_{t+1} + (1 - \iota' x) R_t^f = x' r_{t+1} + R_t^f$ and the mean–variance problem can be expressed in terms of excess returns:

$$\min_x \mathrm{var}[r_p] = x' \Sigma x \quad \text{subject to} \quad \mathrm{E}[r_p] = x' \mu = \overline{\mu}. \tag{2.5}$$

The solution to this problem is much simpler than in the case without a risk-free asset:

$$x^\star = \underbrace{\frac{\overline{\mu}}{\mu' \Sigma^{-1} \mu}}_{\lambda} \times \Sigma^{-1} \mu, \tag{2.6}$$

where λ is a constant that scales proportionately all elements of $\Sigma^{-1}\mu$ to achieve the desired portfolio risk premium $\overline{\mu}$. From this expression, the weights of the tangency portfolio can be found simply by noting that the weights of the tangency portfolio must sum to one, because it lies on the mean–variance frontier of the risky assets. For the tangency portfolio:

$$\lambda_{\mathrm{tgc}} = \frac{1}{\iota' \Sigma^{-1} \mu} \quad \text{and} \quad \overline{\mu}_{\mathrm{tgc}} = \frac{\mu' \Sigma^{-1} \mu}{\iota' \Sigma^{-1} \mu}. \tag{2.7}$$

The formulations (2.1) and (2.2) or (2.5) of the mean–variance problem generate a mapping from a predetermined portfolio risk premium $\overline{\mu}$ to the minimum–variance portfolio weights $x^\star$ and resulting portfolio return volatility $\sqrt{x^{\star\prime} \Sigma x^\star}$. The choice of the desired risk premium, however, depends inherently on the investor's tolerance for risk. To incorporate the investor's optimal trade-off between expected return and risk, the mean–variance problem can be formulated alternatively as the following expected utility maximization:

$$\max_x \mathrm{E}[r_{p,t+1}] - \frac{\gamma}{2} \mathrm{var}[r_{p,t+1}], \tag{2.8}$$

where γ measures the investor's level of relative risk aversion. The solution to this maximization problem is given by Eq. (2.6) with $\lambda = 1/\gamma$, which explicitly links the optimal allocation to the tangency portfolio to the investor's tolerance for risk.

The obvious appeal of the Markowitz paradigm is that it captures the two fundamental aspects of portfolio choice – diversification and the trade-off between expected return and risk – in an analytically tractable and easily extendable framework. This has made it the de-facto standard in the finance profession. Nevertheless, there are several common objections to the Markowitz paradigm. First, the mean–variance problem only represents an expected utility maximization for the special case of quadratic utility, which is a problematic preference specification because it is not monotonically increasing in wealth. For all other utility functions, the mean–variance problem can at best be interpreted as a second-order approximation of expected utility maximization. Second, but related, the mean–variance problem ignores any preferences toward higher-order return moments, in particular toward return skewness and kurtosis. In the context of interpreting the mean–variance problem as a second-order approximation, the third and higher-order terms may be economically nonnegligible. Third, the mean–variance problem is inherently a myopic single-period problem, whereas we think of most investment problems as involving longer horizons with intermediate portfolio rebalancing. Each criticism has prompted numerous extensions of the mean–variance paradigm.[1] However, the most straightforward way to address all these issues, and particularly the third, is to formulate the problem explicitly as an intertemporal expected utility maximization.

2.2. Intertemporal Expected Utility Maximization

2.2.1. Discrete Time Formulation

Consider the portfolio choice at time t of an investor who maximizes the expected utility of wealth at some future date $t + \tau$ by trading in N risky assets and a risk-free asset at times $t, t+1, \ldots, t+\tau-1$. The investor's problem is

$$V(\tau, W_t, z_t) = \max_{\{x_s\}_{s=t}^{t+\tau-1}} \mathrm{E}_t\big[u(W_{t+\tau})\big], \tag{2.9}$$

subject to the budget constraint:

$$W_{s+1} = W_s\big(x_s{}' r_{s+1} + R_s^f\big) \tag{2.10}$$

and having positive wealth each period, $W_s \geq 0$. The function $u(\cdot)$ measures the investor's utility of terminal wealth $W_{t+\tau}$, and the subscript on the expectation denotes that

[1]The majority of extensions deal with incorporating higher-order moments. For example, in Brandt et al. (2005), we propose a fourth-order approximation of expected utility maximization that captures preferences toward skewness and kurtosis. While the optimal portfolio weights cannot be solved for analytically, we provide a simple and efficient numerical procedure. Other work on incorporating higher-order moments include Kraus and Litzenberger (1976), Kane (1982), Simaan (1993), de Athayde and Flores (2004), and Harvey et al. (2004).

the expectation is taken conditional on the information set z_t available at time t. For concreteness, think of z_t as a $K < \infty$ dimensional vector of state variables and assume that $y_t \equiv [r_t, z_t]$ evolves as a first-order Markov process with transition density $f(y_t|y_{t-1})$.[2]

The case $\tau = 1$ corresponds to a static single-period optimization. In general, however, the portfolio choice is a more complicated dynamic multiperiod problem. In choosing at date t the optimal portfolio weights x_t conditional on having wealth W_t and information z_t, the investor takes into account that at every future date s the portfolio weights will be optimally revised conditional on the then available wealth W_s and information z_s.

The function $V(\tau, W_t, z_t)$ denotes the investor's expectation at time t, conditional on the information z_t, of the utility of terminal wealth $W_{t+\tau}$ generated by the current wealth W_t and the sequence of *optimal* portfolio weights $\{x_s^\star\}_{s=t}^{t+\tau-1}$ over the next τ periods. $V(\cdot)$ is called the value function because it represents the value, in units of expected utils, of the portfolio choice problem to the investor. Think of the value function as measuring the quality of the investment opportunities available to the investor. If the current information suggests that investment opportunities are good, meaning, for example, that the sequence of optimal portfolio choices is expected to generate an above average return with below average risk, the current value of the portfolio choice problem to the investor is high. If investment opportunities are poor, the value of the problem is low.

The dynamic nature of the multiperiod portfolio choice is best illustrated by expressing the problem (2.9) as a single-period problem with state-dependent utility $V(\tau - 1, W_{t+1}, z_{t+1})$ of next period's wealth W_{t+1} and information z_{t+1}:

$$\begin{aligned} V(\tau, W_t, z_t) &= \max_{\{x_s\}_{s=t}^{t+\tau-1}} \mathrm{E}_t\big[u\big(W_{t+\tau}\big)\big] \\ &= \max_{x_t} \mathrm{E}_t\left[\max_{\{x_s\}_{s=t+1}^{t+\tau-1}} \mathrm{E}_{t+1}\big[u\big(W_{t+\tau}\big)\big]\right] \qquad (2.11) \\ &= \max_{x_t} \mathrm{E}_t\Big[V\Big(\tau - 1, W_t\big(x_t{}'r_{t+1} + R_t^f\big), z_{t+1}\Big)\Big], \end{aligned}$$

subject to the terminal condition $V(0, W_{t+\tau}, z_{t+\tau}) = u(W_{t+\tau})$. The second equality follows from the law of iterated expectations and the principle of optimality. The third equality uses the definition of the value function as well as the budget constraint. It is important to recognize that the expectation in the third line is taken over the *joint* distribution of next period's returns r_{t+1} and information z_{t+1}, conditional on the current information z_t.

[2] The first-order assumption is innocuous because z_t can contain lagged values.

Equation (2.11) is the so-called Bellman equation and is the basis for any recursive solution of the dynamic portfolio choice problem. The FOCs for an optimum at each date t are[3]

$$\mathrm{E}_t\Big[V_2\big(\tau-1, W_t\big(x_t{}'r_{t+1}+R_t^f\big), z_{t+1}\big)r_{t+1}\Big]=0, \tag{2.12}$$

where $V_i(\cdot)$ denotes the partial derivative with respect to the ith argument of the value function. These FOCs make up a system of nonlinear equations involving possibly high-order integrals and can in general be solved for x_t only numerically.

CRRA Utility Example For illustrative purposes, consider the case of constant relative risk aversion (CRRA) utility $u(W_{t+\tau}) = W_{t+\tau}{}^{1-\gamma}/(1-\gamma)$, where γ denotes the coefficient of relative risk aversion. The Bellman equation then simplifies to:

$$\begin{aligned}
V(\tau, W_t, z_t) &= \max_{x_t} \mathrm{E}_t\left[\max_{\{x_s\}_{s=t+1}^{t+\tau-1}} \mathrm{E}_{t+1}\left[\frac{W_{t+\tau}{}^{1-\gamma}}{1-\gamma}\right]\right]\\
&= \max_{x_t} \mathrm{E}_t\left[\max_{\{x_s\}_{s=t+1}^{t+\tau-1}} \mathrm{E}_{t+1}\left[\frac{\big(W_t\prod_{s=t}^{t+\tau-1}\big(x_s{}'r_{s+1}+R_s^f\big)\big)^{1-\gamma}}{1-\gamma}\right]\right]\\
&= \max_{x_t} \mathrm{E}_t\Bigg[\underbrace{\frac{\big(W_t\big(x_t{}'r_{t+1}+R_t^f\big)\big)^{1-\gamma}}{1-\gamma}}_{u(W_{t+1})}\underbrace{\max_{\{x_s\}_{s=t+1}^{t+\tau-1}} \mathrm{E}_{t+1}\Big[\Big(\prod_{s=t+1}^{t+\tau-1}\big(x_s{}'r_{s+1}+R_s^f\big)\Big)^{1-\gamma}\Big]}_{\psi(\tau-1, z_{t+1})}\Bigg]
\end{aligned} \tag{2.13}$$

In words, with CRRA utility the value function next period, $V(\tau-1, W_{t+1}, z_{t+1})$, is equal to the product of the utility of wealth $u(W_{t+1})$ and a function $\psi(\tau-1, z_{t+1})$ of the horizon $\tau-1$ and the state variables z_t. Furthermore, as the utility function is homothetic in wealth we can, without loss of generality, normalize $W_t = 1$. It follows that the value function depends only on the horizon and state variables, and that the Bellman equation is

$$\frac{1}{1-\gamma}\,\psi(\tau, z_t) = \max_{x_t} \mathrm{E}_t\left[\frac{\big(x_t{}'r_{t+1}+R_t^f\big)^{1-\gamma}}{1-\gamma}\,\psi\big(\tau-1, z_{t+1}\big)\right]. \tag{2.14}$$

The corresponding FOCs are

$$\mathrm{E}_t\Big[\big(x_t{}'r_{t+1}+R_t^f\big)^{-\gamma}\,\psi\big(\tau-1, z_{t+1}\big)r_{t+1}\Big]=0, \tag{2.15}$$

which, despite being simpler than in the general case, can still only be solved numerically.

[3] As long as the utility function is concave, the second-order conditions are satisfied.

The Bellman equation for CRRA utility illustrates how the dynamic and myopic portfolio choices can differ. If the excess returns r_{t+1} are contemporaneously independent of the innovations to the state variables z_{t+1}, the optimal τ and one-period portfolio choices at date t are identical because the conditional expectation in the Bellman equation factors into a product of two conditional expectations. The first expectation is of the utility of next period's wealth $u(W_{t+1})$, and the second is of the function of the state variables $\psi(\tau - 1, z_{t+1})$. Because the latter expectation does not depend on the portfolio weights, the FOCs of the multiperiod problem are the same as those of the single-period problem. If, in contrast, the excess returns are not independent of the innovations to the state variables, the conditional expectation does not factor, the FOCs are not the same, and, as a result, the dynamic portfolio choice may be substantially different from the myopic portfolio choice. The differences between the two policies are called *hedging demands* because by deviating from the single-period portfolio choice the investor tries to hedge against changes in the investment opportunities.

More concretely, consider as data generating process $f(y_t|y_{t-1})$ the following restricted and homoscedastic vector auto-regression (VAR) for the excess market return and dividend yield (in logs):[4]

$$\begin{bmatrix} \ln(1 + r_{t+1}) \\ \ln \mathrm{d}p_{t+1} \end{bmatrix} = \beta_0 + \beta_1 \ln \mathrm{d}p_t + \varepsilon_{t+1}, \tag{2.16}$$

where $\mathrm{d}p_{t+1}$ denotes the dividend-to-price ratio and $\varepsilon_{t+1} \overset{\text{i.i.d.}}{\sim} \mathrm{MVN}[0, \Sigma]$. Table 5.1 presents ordinary least squares (OLS) estimates of this return model for quarterly real data on the value weighted CRSP index and 90-day Treasury bill rates from April 1952

Table 5.1 OLS estimates of the VAR using quarterly real data on the value weighted CRSP index and 90-day Treasury bill rates from April 1952 through December 1996

Dependent variable	Intercept	$\ln \mathrm{d}p_t$	$\mathrm{var}[\varepsilon_{t+1}]$ $(\times 10^{-3})$
$\ln(1 + r_{t+1})$	0.2049	0.0568	
	(0.0839)	(0.0249)	$\begin{bmatrix} 6.225 & -6.044 \\ -6.044 & 6.316 \end{bmatrix}$
$\ln \mathrm{d}p_{t+1}$	−0.1694	0.9514	
	(0.0845)	(0.0251)	

Standard errors in parentheses.

[4]This data generating process is motivated by the evidence of return predictability by the dividend yield (e.g., Campbell and Shiller, 1988; Fama and French, 1988) and has been used extensively in the portfolio choice literature (e.g., Barberis, 2000; Campbell and Viceira, 1999; Kandel and Stambaugh, 1996).

through December 1996.[5] The equation-by-equation adjusted R^2s are 2.3 and 89.3%, reflecting the facts that is it quite difficult to forecast excess returns and that the dividend yield is highly persistent and predictable.

Taking these estimates of the data generating process as the truth, the FOCs (2.15) can be solved numerically using a variety of dynamic programming methods (see Judd, 1998, for a review of numerical methods for dynamic programming). Figure 5.2 presents the solution to the single-period (one-quarter) problem. Plot A shows the optimal fraction of wealth invested in stocks $x_t^\star$ as a function of the dividend yield. Plot B shows the corresponding annualized certainty equivalent rate of return $R_t^{\text{ce}}(\tau)$, defined as the risk-free rate that makes the investor indifferent between holding the optimal portfolio and earning the certainty equivalent rate over the next τ periods.[6] The solid, dashed-dotted, and dotted lines are for relative risk aversion γ of 2, 5, and 10, respectively.

At least three features of the solution to the single-period problem are noteworthy. First, both the optimal allocation to stocks and the certainty equivalent rate increase with the dividend yield, which is consistent with the fact that the equity risk premium increases with the dividend yield. Second, the extent to which the investor tries to time the market decreases with risk aversion. The intuition is simple. When the risk premium increases, stocks become more attractive (higher expected return for the same

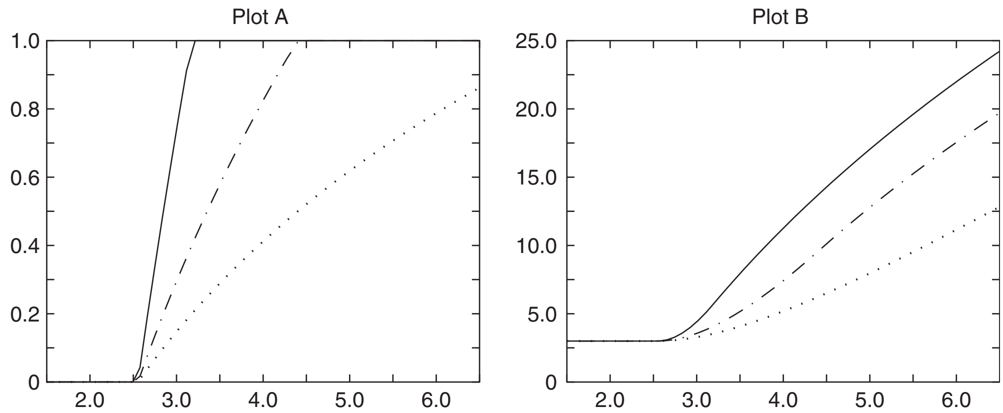

Figure 5.2 Plot A shows the optimal fraction of wealth invested in stocks as a function of the dividend yield for a CRRA investor with one-quarter horizon and relative risk aversion of 2 (solid line), 5 (dashed-dotted line), and 10 (dotted line). Plot B shows the corresponding annualize certainty equivalent rates of return (in percent).

[5]Note that the evidence of return predictability by the dividend yield has significantly weakened over the past 7 years (1997–2003) (e.g., Ang and Bekaert, 2007; Goyal and Welch, 2003). I ignore this most recent sample period for illustrative purposes and to reflect the literature on portfolio choice under return predictability by the dividend yield (e.g., Barberis, 2000; Campbell and Viceira, 1999; Kandel and Stambaugh, 1996). However, keep in mind that the results do not necessarily reflect the current data.

[6]For CRRA utility, the certainty equivalent rate is defined by $\left[R_t^{\text{ce}}(\tau)W_t\right]^{1-\gamma}/(1-\gamma) = V(\tau, W_t, z_t)$.

level of risk), and consequently the investor allocates more wealth to stocks. As the stock allocation increases, the mean of the portfolio return increases linearly while the variance increases quadratically and hence at some point increases faster than the mean. Ignoring higher-order moments, the optimal allocation sets the expected utility gain from a marginal increase in the portfolio mean to equal the expected utility loss from the associated increase in the portfolio variance. The willingness to trade off expected return for risk at the margin depends on the investor's risk aversion. Third, the benefits from market timing also decrease with risk aversion. This is because a more risk averse investor allocates less wealth to stocks and therefore has a lower expected portfolio return and because, even for the same expected portfolio return, a more risk averse investor requires a smaller incentive to abstain from risky investments.

Figure 5.3 presents the solution to the multiperiod portfolio choice for horizons τ ranging from one quarter to 10 years for an investor with $\gamma = 5$ (corresponding to the dashed-dotted lines in Fig. 5.2). Rather than plotting the entire policy fuction for each horizon, plot A shows only the allocations for current dividend yields of 2.9% (25th percentile, dotted line), 3.5% (median, dashed-dotted line), and 4.1% (75th percentile, solid line). Plot B shows the expected utility *gain*, measured by the increase in the annualized certainty equivalent rates (in percent), from implementing the dynamic multiperiod portfolio policy as opposed to making a sequence of myopic single-period portfolio choices.

It is clear from plot A that the optimal portfolio choice depends on the investor's horizon. At the median dividend yield, for example, the optimal allocation is 58% stocks for a one-quarter horizon (one period), 66% stocks for a 1-year horizon (four periods),

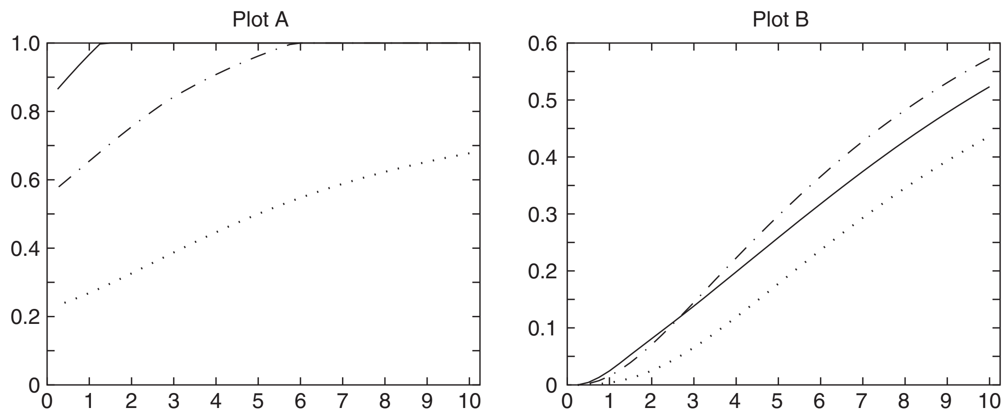

Figure 5.3 Plot A shows the optimal fraction of wealth invested in stocks as a function of the investment horizon for a CRRA investor with relative risk aversion of five conditional on the current dividend yield being equal to 2.9 (dotted line), 3.5 (dashed-dotted line), and 4.1 (solid line) percent. Plot B shows the corresponding increase in the annualized certainty equivalent rates of return from investing optimally as opposed to myopically (in percent).

96% stocks for a 5-year horizon (20 periods), and 100% stocks for all horizons longer than 6 years (24 periods). The differences between the single-period allocations (23, 58, and 87% stocks at the 25th, 50th, and 75th percentiles of the dividend yield, respectively) and the corresponding multiperiod allocations represent the investor's hedging demands. Plot B shows that these hedging demands can lead to substantial increases in expected utility. At the median dividend yield, the increase in the certainty equivalent rate is 2 basis points per year for the 1-year problem, 30 basis points per year for the 5-year problem, and 57 basis points per year for the 10-year problem. Although these gains are small relative to the level of the certainty equivalent rate (5.2% at the median dividend yield), they are large when we ask "how much wealth is the investor willing to give up today to invest optimally, as opposed to myopically, for the remainder of the horizon?" The answer is less than 0.1% for a 1-year investor, but 1.5% for a 5-year investor and 5.9% for a 10-year investor.

Although it is not the most realistic data generating process, the homoscedastic VAR has pedagogical value. First, it demonstrates that in a multiperiod context the optimal portfolio choice can be substantially different from a sequence of single-period portfolio choices, both in terms of allocations and expected utilities. Second, it illustrates the mechanism by which hedging demands arise. The expected return increases with the dividend yield and the higher-order moments are constant. A high (low) dividend yield therefore implies a relatively high (low) value of the portfolio choice problem. In a multiperiod context, this link between the dividend yield and the value of the problem means that the investor faces not only the uncertainty inherent in returns but also uncertainty about whether in the future the dividend yield will be higher, lower, or the same and whether, as a result, the investment opportunities will improve, deteriorate, or remain the same, respectively. Analogous to diversifying cross-sectionally the return risk, the investor wants to smooth intertemporally this risk regarding future investment opportunities. Because the VAR estimates imply a large negative correlation between the stock returns and innovations to the dividend yield, the investment opportunities risk can be smoothed quite effectively by *over-investing* in stocks, relative to the myopic allocation. By over-investing, the investor realizes a greater gain when the return is positive and a greater loss when it is negative. A positive return tends to be associated with a drop in the dividend yield and an expected utility loss due to deteriorated investment opportunities in the future. Likewise, a negative return tends to be associated with a rise in the dividend yield and an expected utility gain due to improved investment opportunities. Thus, the financial gain (loss) from over-investing partially offsets the expected utility loss (gain) associated with the drop (rise) in the dividend yield (hence, the name "hedging demands").

2.2.2. Continuous-Time Formulation

The intertemporal portfolio choice problem can alternatively be expressed in continuous time. The main advantage of the continuous-time formulation is its analytical tractability.

As Merton (1975) and the continuous-time finance literature that followed demonstrates, stochastic calculus allows us to solve in closed-form portfolio choice problems in continuous-time that are analytically intractable in discrete time.[7]

The objective function in the continuous-time formulation is the same as in Eq. (2.9), except that the maximization is over a continuum of portfolio choices x_s, with $t \leq s < t + \tau$, because the portfolio is rebalanced at every instant in time. Assuming that the risky asset prices p_t and the vector of state variables evolve jointly as correlated Itô vector processes:

$$\begin{aligned} \frac{\mathrm{d}p_t}{p_t} - r\mathrm{d}t &= \mu^p(z_t, t)\mathrm{d}t + D^p(z_t, t)\mathrm{d}B_t^p \\ \mathrm{d}z_t &= \mu^z(z_t, t)\mathrm{d}t + D^z(z_t, t)\mathrm{d}B_t^z, \end{aligned} \tag{2.17}$$

the budget constraint is

$$\frac{\mathrm{d}W_t}{W_t} = \left(x_t{}'\mu_t^p + r\right)\mathrm{d}t + x_t{}'D_t^p\mathrm{d}B_t^p, \tag{2.18}$$

Using the abbreviated notation $f_t = f(z_t, t)$, μ_t^p and μ_t^z are N- and K-dimensional conditional mean vectors, D_t^p and D_t^z are $N \times N$ and $K \times K$ conditional diffusion matrices that imply covariance matrices $\Sigma_t^p = D_t^p D_t^{p\prime}$ and $\Sigma_t^z = D_t^z D_t^{z\prime}$, and B_t^p and B_t^z are N- and K-dimensional vector Brownian motion processes with $N \times K$ correlation matrix ρ_t. Finally, r denotes here the instantaneous riskfree rate (assumed constant for notational convenience).

The continuous time Bellman equation is (Merton, 1969):

$$\begin{aligned} 0 = \max_{x_t} \Big[& V_1(\cdot) + W_t \left(x_t{}'\mu_t^p + r\right) V_2(\cdot) + \mu_t^{z\prime} V_3(\cdot) + \frac{1}{2} W_t^2 \, x_t{}' \Sigma_t^p x_t \, V_{22}(\cdot) \\ & + W_t x_t{}' D_t^p \rho_t{}' D_t^{z\prime} V_{23}(\cdot) + \frac{1}{2} \mathrm{tr}\left[\Sigma_t^z V_{33}(\cdot)\right]\Big], \end{aligned} \tag{2.19}$$

subject to the terminal condition $V(0, W_{t+\tau}, z_{t+\tau}) = u(W_{t+\tau})$.

As one might expect, Eq. (2.19) is simply the limit, as $\Delta t \to 0$, of the discrete time Bellman equation (2.11). To fully appreciate this link between the discrete and continuous time formulations, rearrange Eq. (2.11) as:

$$0 = \max_{x_t} \mathrm{E}_t\left[V(\tau - 1, W_{t+1}, z_{t+1}) - V(\tau, W_t, z_t)\right] \tag{2.20}$$

and take the limit of $\Delta t \to 0$:

$$0 = \max_{x_t} \mathrm{E}_t\left[\mathrm{d}V(\tau, W_t, z_t)\right]. \tag{2.21}$$

[7] See Shimko (1999) for an introduction to stochastic calculus. Mathematically more rigorous treatments of the material can be found in Karatzas and Shreve (1991) and Steele (2001).

Then, apply Itô's lemma to the value function to derive:

$$dV(\cdot) = V_1(\cdot)dt + V_2(\cdot)dW_t + V_3(\cdot)dz_t + V_{22}(\cdot)dW_t^2 + V_{23}(\cdot)dW_t dz_t + V_{33}(\cdot)dz_t^2. \tag{2.22}$$

Finally, take the expectation of Eq. (2.22), which picks up the drifts of dW_t, dz_t, $dW_t^2, dW_t\, dz_t$, and dz_t^2 (the second-order processes must be derived through Itô's lemma), plug it into Eq. (2.21), and cancel out the common term dt. The result is Eq. (2.19).

The continuous-time FOCs are

$$\mu_t^p\, V_2(\cdot) + W_t\, x_t{}'\Sigma_t^p\, V_{22}(\cdot) + D_t^p \rho_t{}'\, D^{z\prime} V_{23} = 0, \tag{2.23}$$

which we can solve for the optimal portfolio weights:

$$x_t^\star = \underbrace{-\frac{V_2(\cdot)}{W_t\, V_{22}(\cdot)} \left(\Sigma_t^p\right)^{-1} \mu_t^p}_{\text{myopic demand}} \; \underbrace{- \frac{V_2(\cdot)}{W_t\, V_{22}(\cdot)} \frac{V_{23}(\cdot)}{V_2(\cdot)} (\Sigma_t^p)^{-1}\, D_t^p \rho_t{}' D_t^{z\prime}}_{\text{hedging demand}}. \tag{2.24}$$

This analytical solution illustrates more clearly the difference between the dynamic and myopic portfolio choice. The optimal portfolio weights $x_t^\star$ are the sum of two terms, the first being the myopically optimal portfolio weights and the second representing the difference between the dynamic and myopic solutions. Specifically, the first term depends on the ratio of the first to second moments of excess returns and on the inverse of the investor's relative risk aversion $\gamma_t \equiv -W_t V_{22}(\cdot)/V_2(\cdot)$. It corresponds to holding a fraction $1/\gamma_t$ in the tangency portfolio of the instantaneous mean–variance frontier. The second term depends on the *projection* of the state variable innovations dB_t^z onto the return innovations dB_t^p, which is given by $\left(\Sigma_t^p\right)^{-1} D_t^p \rho_t{}' D_t^{z\prime}$, on the inverse of the investor's relative risk aversion, and on the sensitivity of the investor's marginal utility to the state variables $V_{23}(\cdot)/V_2(\cdot)$. The projection delivers the weights of K portfolios that are *maximally correlated* with the state variable innovations and the derivatives of marginal utility with respect to the state variables measure how important each of these state variables is to the investor. Intuitively, the investor takes positions in each of the maximally correlated portfolios to partially hedge against undesirable innovations in the state variables. The maximally correlated portfolios are therefore called *hedging portfolios*, and the second term in the optimal portfolio weights is labeled the hedging demand. It is important to note that both the myopic and hedging demands are scaled equally by relative risk aversion and that the trade-off between holding a myopically optimal portfolio and intertemporal hedging is determined by the derivatives of marginal utility with respect to the state variables.

CRRA Utility Example Continued To illustrate the tractability of the continuous-time formulation, consider again the CRRA utility example. *Conjecture* that the value function has the separable form:

$$V(\tau, W_t, z_t) = \frac{W_t^{1-\gamma}}{1-\gamma}\,\psi(\tau, z_t), \tag{2.25}$$

which implies that the optimal portfolio weights are

$$x_t^{\star} = \frac{1}{\gamma}\big(\Sigma_t^p\big)^{-1}\,\mu_t^p + \frac{1}{\gamma}\frac{\psi_2(\cdot)}{\psi(\cdot)}\,\big(\Sigma_t^p\big)^{-1} D_t^p \rho_t' D_t^{z\prime}. \tag{2.26}$$

This solution is sensible given the well-known properties of CRRA utility. Both the tangency and hedging portfolio weights are scaled by a constant $1/\gamma$ and the relative importance of intertemporal hedging, given by $\psi_2(\cdot)/\psi(\cdot)$, is independent of wealth.

Plugging the derivatives of the value function (2.25) and the optimal portfolio weights (2.26) into the Bellman equation (2.19), yields the nonlinear differential equation:

$$\begin{aligned} 0 = {} & \psi_1(\cdot) + (1-\gamma)\big(x_t^{\star\prime}\mu_t^p + r\big)\,\psi(\cdot) + \mu_t^{z\prime}\,\psi_2(\cdot) - \frac{1}{2}\,\gamma(1-\gamma)\,x_t^{\star\prime}\,\Sigma_t^p x_t^{\star}\psi(\cdot) \\ & + (1-\gamma)x_t^{\star\prime}\,D_t^p\rho_t'\,D_t^z\,\psi_2(\cdot) + \frac{1}{2}\mathrm{tr}\big[\Sigma_t^z\,\psi_{22}(\cdot)\big]. \end{aligned} \tag{2.27}$$

The fact that this equation, which implicitly defines the function $\psi(\tau, z_t)$, does not depend on the investor's wealth W_t confirms the conjecture of the separable value function.

Continuous Time Portfolio Policies in Discrete Time Because the continuous-time Bellman equation is the limit of its discrete-time counterpart, it is tempting to think that the *solutions* to the two problems share the same limiting property. Unfortunately, this presumption is wrong. The reason is that the continuous time portfolio policies are often *inadmissible* in discrete time because they cannot guarantee nonnegative wealth unless the portfolio is rebalanced at every instant.

Consider a simpler example of logarithmic preferences (CRRA utility with $\gamma = 1$) and i.i.d. log-normal stock returns with annualized risk premium of 5.7% and volatility of 16.1% (consistent with the VAR in the previous section). In the continuous-time formulation, the optimal stock allocation is $x_t^{\star} = 0.057/0.161^2 = 2.20$, which means that the investor borrows 120% of wealth to invest a total of 220% in stocks. Technically, such levered position is inadmissable over *any* discrete time interval, irrespective of how short it is. The reason is that under log-normality the gross return on stocks over any finite interval can be arbitrarily close to zero, implying a positive probability that the investor cannot repay the loan next period. This constitutes a possible violation of the no-bankruptcy constraint $W_s \geq 0$ and, with CRRA utility, can lead to infinite disutility. The continuous-time solution is therefore inadmissable in discrete time, and the optimal discrete-time allocation is $x_t^\star \leq 1$.

Whether this inadmissability is important enough to abandon the analytical convenience of the continuous-time formulation is up to the researcher to decide. On the one hand, the probability of bankruptcy is often very small. In the log utility example, for instance, the probability of realizing a sufficiently negative stock return over the period of one quarter is only 1.3×10^{-9}. On the other hand, in reality an investor always faces some risk of loosing all, or almost all wealth invested in risky securities due to an extremely rare but severe event, such as a stock market crash, the collapse of the financial system, or investor fraud.[8]

2.3. When is it Optimal to Invest Myopically?

Armed with the discrete and continuous-time formulations of the portfolio choice problem, we can be more explicit about when it is optimal to invest myopically. The myopic portfolio choice is an important special case for practitioners and academics alike. There are, to my knowledge, few financial institutions that implement multi-period investment strategies involving hedging demands.[9] Furthermore, until recently the empirically oriented academic literature on portfolio choice was focused almost exclusively on single-period problems, in particular, the mean–variance paradigm of Markowitz (1952) discussed in Section 2.1.

In addition to the obvious case of having a single-period horizon, it is optimal to invest myopically under each of the following three assumptions:

2.3.1. Constant Investment Opportuntities

Hedging demands only arise when the investment opportunities vary stochastically through time. With constant investment opportunities, the value function does not depend on the state variables, so that z_t drops out of the discrete time FOCs (2.12) and $V_{23}(\cdot) = 0$ in the continuous-time solution (2.24). The obvious case of constant investment opportunities is i.i.d. returns. However, the investment opportunities can be constant even when the conditional moments of returns are stochastic. For example, Nielsen and Vassalou (2006) show that in the context of the diffusion model (2.17), the investment opportunities are constant as long as the instantaneous riskfree rate and the Sharpe ratio of the optimal portfolio of an investor with logarithmic preferences are

[8] Guided by this rare events argument, there are at least two ways to formally bridge the gap between the discrete and continuous-time solutions. We can either introduce the rare events through jumps in the continuous-time formulation (e.g., Longstaff et al., 2003) or allow the investor to purchase insurance against the rare events through put options or other derivatives in the discrete-time formulation.

[9] A common justification from practitioners is that the expected utility loss from errors that could creep into the solution of a complicated dynamic optimization problem outweighs the expected utility gain from investing optimally as opposed to myopically. Recall that in the dividend yield predictability case the gain for CRRA utility is only a few basis points per year.

constant. The conditional means, variances, and covariances of the individual assets that make up this log-optimal portfolio can vary stochastically.

2.3.2. Stochastic but Unhedgable Investment Opportunities

Even with stochastically varying investment opportunities, hedging demands only arise when the investor can use the available assets to hedge against changes in future investment opportunities. If the variation is completely independent of the returns, the optimal portfolio is again myopic. In discrete time, independence of the state variables and returns implies that the expectation in the Bellman equation can be decomposed into an expectation with respect to the portfolio returns and an expectation with respect to the state variables. The FOCs then turn out to be the same as in the single-period problem. In continuous time, a correlation $\rho_t = 0$ between the return and state variable innovations eliminates the hedging demands term in the optimal portfolio weights.

2.3.3. Logarithmic Utility

Finally, the portfolio choice reduces to a myopic problem when the investor has logarithmic preferences $u(W) = \ln(W)$. The reason is that with logarithmic preferences the utility of terminal wealth is simply the sum of the utilities of single-period portfolio returns:

$$\ln(W_{t+\tau}) = \ln\left(W_t \prod_{s=t}^{t+\tau-1} \left(x_s' r_{s+1} + R_s^f\right)\right) = \ln W_t + \sum_{s=t}^{t+\tau-1} \ln\left(x_s' r_{s+1} + R_s^f\right). \tag{2.28}$$

The portfolio weights that maximize the expectation of the sum are the same as the ones that maximize the expectations of each element of the sum, which are, by definition, the sequence of single-period portfolio weights. Therefore, the portfolio choice is myopic.

2.4. Modeling Issues and Extensions

2.4.1. Preferences

The most critical ingredient to any portfolio choice problem is the objective function. Historically, the academic literature has focused mostly on time-separable expected utility with hyperbolic absolute risk aversion (HARA), which includes as special cases logarithmic utility, power or constant relative risk aversion (CRRA) utility, negative exponential or constant absolute risk aversion (CARA) utility, and quadratic utility. The reason for this popularity is the fact that HARA is a necessary and sufficient condition to obtain asset demand functions expressed in currency units, not percent of wealth, that are linear in wealth (Merton, 1969). In particular, the portfolio choice expressed in currency units is proportional to wealth with CRRA utility and independent of wealth with CARA utility. Alternatively, the corresponding portfolio choice expressed in percent of wealth is independent of wealth with CRRA utility and inversely proportional to wealth with CARA utility.

In the HARA class, power or CRRA preferences are by far the most popular because the value function turns out to be homogeneous in wealth (see the examples mentioned earlier). However, CRRA preferences are not without faults. One critique that is particularly relevant in the portfolio choice context is that with CRRA the elasticity of intertemporal substitution is directly tied to the level of relative risk aversion (one is the reciprocal of the other), which creates an unnatural link between two very different aspects of the investor's preferences – the willingness to substitute consumption intertemporally versus the willingness to take on risk. Epstein and Zin (1989) and Weil (1989) propose a generalization of CRRA preferences based on recursive utility that severs this link between intertemporal substitution and risk aversion. Campbell and Viceira (1999) and Schroder and Skiadas (1999) consider these generalized CRRA preferences in portfolio choice problems.

A number of stylized facts of actual investment decisions and professional investment advice are difficult to reconcile with HARA or even Epstein–Zin–Weil preferences. The most prominent empirical anomaly is the strong dependence of observed and recommended asset allocations on the investment horizon.[10] There have been a number of attempts to explain this horizon puzzle using preferences in which utility is defined with respect to a nonzero and potentially time-varying lower bound on wealth or consumption, including a constant subsistence level (Jagannathan and Kocherlakota, 1996; Samuelson, 1989), consumption racheting (Dybvig, 1995), and habit formation (Lax, 2002; Schroder and Skiadas, 2002).

Experiments by psychologists, sociologists, and behavioral economists have uncovered a variety of more fundamental behavioral anomalies. For example, the way experimental subjects make decisions under uncertainty tends to systematically violate the axioms of expected utility theory (e.g., Camerer, 1995). To capture these behavioral anomalies in an optimizing framework, several nonexpected utility preference formulations have been proposed, including loss aversion and prospect theory (Kahneman and Tversky, 1979), anticipated or rank-dependent utility (Quiggin, 1982), ambiguity aversion (Gilboa and Schmeidler, 1989), and disappointment aversion (Gul, 1991). These nonexpected utility preferences have been applied to portfolio choice problems by Benartzi and Thaler (1995), Shefrin and Statman (2000), Aït-Sahalia and Brandt (2001), Liu (2002), Ang et al. (2005), and Gomes (2005), among others.

Finally, there are numerous applications of more practitioner-oriented objective functions, such as minimizing the probability of a short-fall (Kataoka, 1963; Roy, 1952; Telser, 1956), maximizing expected utility with either absolute or relative portfolio insurance (Black and Jones, 1987; Grossman and Vila, 1989; Perold and Sharpe, 1988), maximizing expected utility subject to beating a stochastic benchmark (Browne, 1999; Tepla, 2001),

[10] E.g., see Bodie and Crane (1997), Canner et al. (1997), and Ameriks and Zeldes (2004).

and maximizing expected utility subject to maintaining a critical value at risk (VaR) (Alexander and Baptista, 2002; Basak and Shapiro, 2001; Cuoco et al., 2007).

2.4.2. Intermediate Consumption

Both the discrete- and continuous-time formulations of the portfolio choice problem can be amended to accommodate intermediate consumption. Simply add to the utility of terminal wealth (interpreted then as the utility of bequests to future generations) the utility of the life-time consumption stream (typically assumed to be time-separable and geometrically discounted), and replace in the budget constraint the current wealth W_t with the current wealth net of consumption $(1-c_t)W_t$, where c_t denotes the fraction of wealth consumed. The investor's problem with intermediate consumption then is to choose at each date t the optimal consumption c_t as well as the asset allocation x_t.

For example, the discrete-time problem with time-separable CRRA utility of consumption and without bequests is

$$V(\tau, W_t, z_t) = \max_{\{x_s, c_s\}_{s=t}^{t+\tau-1}} \mathrm{E}_t\left[\sum_{s=t}^{t+\tau} \beta^{s-t} \frac{(c_t W_t)^{1-\gamma}}{1-\gamma}\right], \tag{2.29}$$

subject to the budget constraint:

$$W_{s+1} = (1-c_s)W_s\left(x_s{}' r_{s+1} + R_s^f\right), \tag{2.30}$$

the no-bankruptcy constraint $W_s \geq 0$, and the terminal condition $c_{t+\tau} = 1$. Following a few steps analogous to the case without intermediate consumption, the Bellman equation can in this case be written as:

$$\frac{1}{1-\gamma}\psi(\tau, z_t) = \max_{x_t, c_t}\left[\frac{c_t^{1-\gamma}}{1-\gamma} + \beta\,\mathrm{E}_t\left[\frac{\left((1-c_t)\left(x_t{}' r_{t+1} + R_t^f\right)\right)^{1-\gamma}}{1-\gamma}\psi(\tau-1, z_{t+1})\right]\right], \tag{2.31}$$

where $\psi(\tau, z_t)$ is again a function of the horizon and state variables that is in general different from the case without intermediate consumption.

Although the Bellman equation with intermediate consumption is more involved than without, in the case of CRRA utility the problem is actually easier to handle numerically because the value function can be solved for explicitly from the envelope condition $\partial V(\tau, W, z)/\partial W = \partial u(cW)/\partial(cW)$. Specifically, $\psi(\tau, z) = c(\tau, z)^{-\gamma}$ for $\gamma > 0$ and $\gamma \neq 1$ or $\psi(\tau, z) = 1$ for $\gamma = 1$. This explicit form of the value function implies that in a backward-recursive dynamic programming solution to the policy functions $x(\tau, z)$ and $c(\tau, z)$, the value function at date $t+1$, which enters the FOCs at date t, is automatically provided by the consumption policy at date $t+1$ obtained in the previous recursion. Furthermore, with CRRA utility the portfolio and consumption choices turn out to be sequential. Because the value function is homothetic in wealth and the consumption choice c_t only scales the investable wealth $(1 - c_t)W_t$, the FOCs for the portfolio weights x_t are independent of c_t. Therefore, the investor first makes the portfolio choice ignoring consumption and then makes the consumption choice given the optimal portfolio weights.

As Wachter (2002) demonstrates, the economic implication of introducing intermediate consumption in a CRRA framework is to shorten the effective horizon of the investor. Although the myopic portfolio choice is the same with and without intermediate consumption, the hedging demands are quite different in the two cases. In particular, Wachter shows that the hedging demands with intermediate consumption are a weighted sum of the hedging demands of a sequence of terminal wealth problems, analogous to the price of a coupon-bearing bond being a weighted sum of the prices of a sequence of zero-coupon bonds.

2.4.3. Complete Markets

A financial market is said to be complete when all future outcomes (states) are spanned by the payoffs of traded assets. In a complete market, state-contingent claims or so-called Arrow–Debreu securities that pay off one unit of consumption in a particular state and zero in all other states can be constructed for every state. These state-contingent claims can then be used by investors to place bets on a particular state or set of states.

Markets can be either statically or dynamically complete. For a market to be statically complete, there must be as many traded assets as there are states, such that investors can form state-contingent claims as buy-and-hold portfolios of these assets. Real asset markets, in which there is a continuum of states and only a finite number of traded assets, are at best dynamically complete. In a dynamically complete market, investors can construct a continuum of state-contingent claims by *dynamically* trading in the finite set of base assets. Dynamic completion underlies, for example, the famous Black and Scholes (1973) model and the extensive literature on derivatives pricing that followed.[11]

The assumption of complete markets simplifies not only the pricing of derivatives but, as Cox and Huang (1989, 1991) demonstrate, also the dynamic portfolio choice. Rather than solve for a dynamic trading strategy in a set of base assets, Cox and Huang solve for the optimal buy-and-hold portfolio of the state-contingent claims. The intuition is that any dynamic trading strategy in the base assets generates a particular terminal payoff distribution that can be replicated by some buy-and-hold portfolio of state-contingent claims. Conversely, any state-contingent claim can be replicated by a dynamic trading

[11] Dynamic completion arises usually in a continuous time setting, but Cox et al. (1979) illustrate that continuous trading is not a critical assumption. They construct an $(N+1)$ state discrete time economy as a sequence of N binomial economies and show that this statically incomplete economy can be dynamically completed by trading in only two assets.

strategy in the base assets. It follows that the terminal payoff distribution generated by the optimal dynamic trading strategy in the base assets is *identical* to that of the optimal static buy-and-hold portfolio of state-contingent claims. Once this static problem is solved (which is obviously much easier than solving the dynamic optimization), the optimal dynamic trading strategy in the base assets can be recovered by adding up the replicating trading strategies of each state-contingent claim position in the buy-and-hold portfolio.

The Cox and Huang (1989, 1991) approach to portfolio choice relies on the existence of a state price density or equivalent Martingale measure (see Harrison and Kreps, 1979) and is therefore often referred to as the "Martingale approach" to portfolio choice. Cox and Huang solve the continuous time HARA problem with intermediate consumption and confirm that the results are identical to the dynamic programming solution of Merton (1969). Recent applications of the Martingale approach to portfolio choice problems with frictionless markets and the usual utility functions include Wachter (2002), who specializes Cox and Huang's solution to CRRA utility and a return process similar to the VAR mentioned earlier, Detemple et al. (2003), who show how to recover the optimal trading strategy in the base assets as opposed to the Arrow–Debreu securities for a more general return processes using simulations, and Aït-Sahalia and Brandt (2007), who incorporate the information in option-implied state prices in the portfolio choice problem.

Although originally intended for solving portfolio choice problems in complete markets, the main success of the Martingale approach has been in the context of problems with incompleteness due to portfolio constraints, transaction costs, and other frictions, which are notoriously difficult to solve using dynamic programming techniques. He and Pearson (1991) explain how to deal with market incompleteness in the Martingale approach. Cvitanic (2001) surveys the extensive literature that applies the Martingale approach to portfolio choice problems with different forms of frictions. Another popular use of the Martingale approach is in the context of less standard preferences (see the references in Section 2.4.1).

2.4.4. Infinite or Random Horizon

Solving an infinite horizon problem is often easier than solving an otherwise identical finite horizon problem because the infinite horizon assumption eliminates the dependence of the Bellman equation on time. An infinite horizon problem only needs to be solved for a steady-state policy, whereas a finite horizon problem must be solved for a different policy each period. For example, Campbell and Viceira (1999) and Campbell et al. (2003) are able to derive approximate analytical solutions to the infinite horizon portfolio choice of an investor with recursive Epstein–Zin–Weil utility, intermediate consumption, and mean-reverting expected returns. The same problem with a finite horizon can only be solved numerically, which is difficult (in particular in the multi-asset case considered by Campbell et al.) and the results are not as transparent as an analytical solution.

Intuitively, one would expect the sequence of solutions to a finite horizon problem to converge to that of the corresponding infinite horizon problem as the horizon increases.[12] In the case of CRRA utility and empirically sensible return processes, this convergence appears to be quite fast. Brandt (1999), Barberis (2000), and Wachter (2002) document that 10- to 15-year CRRA portfolio policies are very similar to their infinite horizon counterparts. This rapid convergence suggests that the solution to the infinite horizon problem can, in many cases, be confidently used to study the properties of long- but finite-horizon portfolio choice in general (e.g., Campbell and Viceira, 1999, 2002).

Having a known finite or an infinite horizon are pedagogical extremes. In reality, an investor rarely knows the terminal date of an investment, which introduces another source of uncertainty. In the case of intermediate consumption, the effect of horizon uncertainty can be substantial because the investor risks either running out of wealth before the terminal date or leaving behind accidental bequests (e.g., Barro and Friedman, 1977; Hakansson, 1969). An alternative motivation for a random terminal date is to set a finite *expected horizon* in an infinite horizon problem to sharpen the approximation of a long-horizon portfolio choice by its easier-to-solve infinite horizon counterpart (e.g., Viceira, 2001).

2.4.5. Frictions and Background Risks

Arguably the two most realistic features of an investor's problem are frictions, such as transaction costs and taxation, and background risks, which refers to any risks other than those directly associated with the risky securities. Frictions are particularly difficult to incorporate because they generally introduce path dependencies in the solution to the portfolio choice problem. For example, with proportional transaction costs, the costs incurred by rebalancing depend on both the desired allocations for the next period and the current allocation inherited from the previous period. In the case of capital gains taxes, the basis for calculating the tax liability generated by selling an asset depends on the price at which the asset was originally bought. Unfortunately, in the usual backward recursive solution of the dynamic program, the previous investment decisions are unknown.

Because of its practical relevance, the work on incorporating frictions, transaction costs and taxation in particular, into portfolio choice problems is extensive and ongoing. Recent papers on transaction costs include Davis and Norman (1990), Duffie and Sun (1990), Akian et al. (1996), Balduzzi and Lynch (1999), Leland (2001), Liu (2004), and Lynch and Tan (2009). The implications of capital gains taxation are considered in a single-period context by Elton and Gruber (1978) and Balcer and Judd (1987) and in a multiperiod context by Dammon et al. (2001a,b), Garlappi et al. (2001), Leland (2001),

[12] Merton (1969) proves this intuition for the continuous time portfolio choice with CRRA utility. Kim and Omberg (1996) provide counter-examples with HARA utility for which the investment problem becomes ill-defined at sufficiently long horizons (so-called nirvana solutions).

Dammon et al. (2004), DeMiguel and Uppal (2005), Gallmeyer et al. (2006), and Huang (2008), among others.

In principle, background risks encompass all risks faced by an investor other than those directly associated with the risky securities. The two most common sources of background risk considered in the academic literature are uncertain labor or entrepreneurial income and both the investment in and consumption of housing. Recent work on incorporating uncertain labor or entrepreneurial income include Heaton and Lucas (1997), Koo (1998), Chan and Viceira (2000), Heaton and Lucas (2000), Viceira (2001), and Gomes and Michaelides (2003). The role of housing in portfolio choice problems is studied by Grossman and Laroque (1991), Flavin and Yamashita (2002), Cocco (2000, 2005), Campbell and Cocco (2003), Hu (2005), and Yao and Zhang (2005), among others. The main challenge in incorporating background risks is to specify a realistic model for the joint distribution of these risks with asset returns at different horizons and over the investor's life-cycle.

3. TRADITIONAL ECONOMETRIC APPROACHES

The traditional role of econometrics in portfolio choice problems is to specify the data generating process $f(y_t|y_{t-1})$. As straightforward as this seems, there are two different econometric approaches to portfolio choice problems: *plug-in estimation* and *decision theory*. In the plug-in estimation approach, the econometrician draws inferences about some investor's optimal portfolio weights to make descriptive statements, while in the decision theory approach, the econometrician takes on the role of the investor and draws inferences about the return distribution to choose portfolio weights that are optimal with respect to these inferences.

3.1. Plug-In Estimation

The majority of the portfolio choice literature, and much of what practitioners do, falls under the heading of plug-in estimation or calibration, where the econometrician estimates or otherwise specifies the parameters of the data generating process and then plugs these parameter values into an analytical or numerical solution to the investor's optimization problem. Depending on whether the econometrician treats the parameters as estimates or simply assumes them to be the truth, the resulting portfolio weights are estimated or calibrated. Estimated portfolio weights inherit the estimation error of the parameter estimates and therefore are almost certainly different from the true optimal portfolio weights in finite samples.

3.1.1. Theory

Single-Period Portfolio Choice Consider first a single-period portfolio choice problem. The solution of the investor's expected utility maximization maps the preference parameters ϕ (e.g., the risk aversion coefficient γ for CRRA utility), the state vector z_t, and the parameters of the data generating process θ into the optimal portfolio weights x_t:

$$x_t^\star = x(\phi, z_t, \theta), \tag{3.1}$$

where ϕ is specified ex-ante and z_t is observed. Given data $Y_T \equiv \{y_t\}_{t=0}^T$, we can typically obtain unbiased or at least consistent estimates $\hat{\theta}$ of the parameters θ. Plugging these estimate into Eq. (3.1) yields estimates of the optimal portfolio weights $\hat{x}_t^\star = x(\phi, z_t, \hat{\theta})$.

Assuming $\hat{\theta}$ is consistent with asymptotic distribution $\sqrt{T}(\hat{\theta} - \theta) \overset{T\to\infty}{\sim} \mathrm{N}[0, V_\theta]$ and the mapping $x(\cdot)$ is sufficiently well-behaved in θ, the asymptotic distribution of the estimator $\hat{x}_t^\star$ can be computed using the delta method:

$$\sqrt{T}\big(\hat{x}_t^\star - x_t^\star\big) \overset{T\to\infty}{\sim} \mathrm{N}\big[0, x_3(\cdot)\, V_\theta x_3(\cdot)'\big]. \tag{3.2}$$

To be more concrete, consider the mean–variance problem (2.8). Assuming i.i.d. excess returns with constant risk premia μ and covariance matrix Σ, the optimal portfolio weights are $x^\star = (1/\gamma)\, \Sigma^{-1}\mu$. Given excess return data $\{r_{t+1}\}_{t=1}^T$, the moments μ and Σ can be estimated using the following sample analog:

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} r_{t+1} \quad \text{and} \quad \hat{\Sigma} = \frac{1}{T-N-2}\sum_{t=1}^{T}(r_{t+1} - \hat{\mu})(r_{t+1} - \hat{\mu})' \tag{3.3}$$

(notice the unusual degrees of freedom of $\hat{\Sigma}$). Plugging these estimates into the expression for the optimal portfolio weights gives the plug-in estimates $\hat{x}^\star = (1/\gamma)\, \hat{\Sigma}^{-1}\hat{\mu}$.

Under the assumption of normality, this estimator is unbiased:

$$\mathrm{E}[\hat{x}^\star] = \frac{1}{\gamma}\,\mathrm{E}[\hat{\Sigma}^{-1}]\,\mathrm{E}[\hat{\mu}] = \frac{1}{\gamma}\,\Sigma^{-1}\mu, \tag{3.4}$$

where the first equality follows from the standard independence of $\hat{\mu}$ and $\hat{\Sigma}$, and the second equality is due to the unbiasedness of $\hat{\mu}$ and $\hat{\Sigma}^{-1}$.[13] Without normality or with the more standard $1/T$ or $1/(T-1)$ normalization for the sample covariance matrix, the plug-estimator is generally biased but nonetheless consistent with plim $\hat{x}^\star = x^\star$.

The second moments of the plug-in estimator can be derived by expanding the estimator around the true risk premia and return covariance matrix. With multiple risky assets, this expansion is algebraicly tedious because of the nonlinearities from the inverse

[13]The unbiasedness of $\hat{\mu}$ is standard. For the unbiasedness of $\hat{\Sigma}^{-1}$, recall that with normality, the matrix $\hat{S} = \sum_{t=1}^T (r_{t+1} - \hat{\mu})(r_{t+1} - \hat{\mu})'$ has a Wishart distribution (the multivariate extension of a chi-squared distribution) with a mean of $(T-1)\Sigma$. Its inverse $\hat{S}^{-1}$ therefore has an inverse Wishart distribution, which has a mean of $(T-N-2)\Sigma^{-1}$ (see Marx and Hocking, 1977). This implies that $\hat{\Sigma}^{-1}$ is an unbiased estimator of Σ^{-1} and explains the unusual degrees of freedom.

of the covariance matrix (see Jobson and Korkie, 1980). To illustrate the technique, consider therefore a single risky asset. Expanding $\hat{x}^\star = (1/\gamma)\hat{\mu}/\hat{\sigma}^2$ around both μ and σ^2 yields:

$$\hat{x}^\star = \frac{1}{\gamma}\frac{1}{\sigma^2}(\mu - \hat{\mu}) - \frac{1}{\gamma}\frac{\mu}{\sigma^4}(\sigma^2 - \hat{\sigma}^2). \tag{3.5}$$

Take variances and rearrange:

$$\text{var}\big[\hat{x}^\star\big] = \frac{1}{\gamma^2}\left(\frac{\mu}{\sigma^2}\right)^2\left(\frac{\text{var}[\hat{\mu}]}{\mu^2} + \frac{\text{var}[\hat{\sigma}^2]}{\sigma^4}\right). \tag{3.6}$$

This expression shows that the imprecision of the plug-in estimator is scaled by the magnitude of the optimal portfolio weight $x^\star = (1/\gamma)\mu/\sigma^2$ and depends on both the imprecision of the risk premia and volatility estimates, each scaled by their respective magnitudes.

To get a quantitative sense for the estimation error, evaluate Eq. (3.6) for some realistic values for μ, σ, var$[\hat{\mu}]$, and var$[\hat{\sigma}^2]$. Suppose, for example, we have 10 years of monthly data on a stock with $\mu = 6\%$ and $\sigma = 15\%$. With i.i.d. data, the standard error of the sample mean is std$[\hat{\mu}] = \sigma/\sqrt{T} = 1.4\%$. Second moments are generally thought of as being more precisely estimated than first moments. Consistent with this intuition, the standard error of the sample variance under i.i.d. normality is std$[\hat{\sigma}^2] = \sqrt{2}\sigma^2/\sqrt{T} = 0.3\%$. Putting together the pieces, the standard error of the plug-in estimator $\hat{x}^\star$ for a reasonable risk aversion of $\gamma = 5$ is equal to 14%, which is large relative to the magnitude of the true $x^\star = 53.3\%$. This example illustrates a more general point: portfolio weights tend to be very imprecisely estimated because the inputs to the estimator are difficult to pin down.

It is tempting to conclude from this example that, at least for the asymptotics, uncertainty about second moments is swamped by uncertainty about first moments. As Cho (2007) illustrates, however, this conclusion hinges critically on the assumption of i.i.d. normality. In particular, the precision of the sample variance depends on the kurtosis of the data. The fatter are the tails, the more difficult it is to estimate second moments because outliers greatly affect the estimates. This means that conditional heteroskedasticity, in particular, can considerably inflate the asymptotic variance of the unconditional sample variance. Returning to the example, suppose that, instead of i.i.d. normality, the conditional variance h_t of returns follows a standard GARCH(1,1) process:

$$h_t = \omega + \alpha\epsilon_{t-1}^2 + \beta h_{t-1}. \tag{3.7}$$

In this case, the variance of the unconditional sample variance is

$$\text{var}\big[\hat{\sigma}^2\big] = \frac{2\sigma^4}{T}\left(1 + \frac{\kappa}{2}\right)\left(1 + \frac{2\rho}{1 - \alpha - \beta}\right), \tag{3.8}$$

where κ denotes the unconditional excess kurtosis of returns and ρ denotes the first-order autocorrelation of the squared return innovations. Both κ and ρ can be computed from the GARCH parameters α and β. With reasonable GARCH parameter values of $\alpha = 0.0175$ and $\beta = 0.9811$, the variance of the sample variance is inflated by a factor of 233.3. As a result, the standard error of $\hat{x}^\star$ is 105.8%, as compared to 14% under i.i.d. normality. Although this example is admittedly extreme (as volatility is close to being nonstationary), it illustrates the point that both return moments, as well as high-order moments for other preferences, can contribute to the asymptotic imprecision of plug-in portfolio weight estimates.

Returning to the computationally more involved case of multiple risky assets, Britten-Jones (1999) derives a convenient way to draw asymptotic inferences about mean–variance optimal portfolio weights. He shows that the plug-in estimates of the tangency portfolio:

$$\hat{x}^\star_{\text{tgc}} = \frac{\hat{\Sigma}^{-1}\hat{\mu}}{\iota'\hat{\Sigma}^{-1}\hat{\mu}} \tag{3.9}$$

can be computed from OLS estimates of the slope coefficients b of regressing a vector of ones on the matrix of excess returns (without intercept):

$$1 = b\, r_{t+1} + u_{t+1}, \tag{3.10}$$

where $\hat{x}^\star_{\text{tgc}} = \hat{b}/(\iota'\hat{b})$. We can therefore use standard OLS distribution theory for $\hat{b}$ to draw inferences about $x^\star_{\text{tgc}}$. For example, testing whether the weight of the tangency portfolio on a particular asset equals zero is equivalent to testing whether the corresponding element of b is zero, which corresponds to a standard t test. Similarly, testing whether an element of $x^\star_{\text{tgc}}$ equals a constant c is equivalent to testing whether the corresponding element of b equals $c(\iota' b)$, which is a linear restriction that can be tested using a joint F test.

Multiperiod Portfolio Choice The discussion mentioned earlier applies directly to both analytical and approximate solutions of multiperiod portfolio choice problems, in which the optimal portfolio weights at time t are functions of the preference parameters ϕ, the state vector z_t, the parameters of the data generating process θ, and perhaps the investment horizon $T - t$. In the case of a recursive numerical solution, however, the portfolio weights at time t depend explicitly on the value function at time $t+1$, which in turn depends on the sequence of optimal portfolio weights at times $\{t+1, t+2, \ldots, T-1\}$. Therefore, the portfolio weight estimates at time t not only reflect the imprecision of the parameter estimates but also the imprecision of the estimated portfolio weights for future periods (which themselves reflect the imprecision of the parameter estimates). To capture this recursive dependence of the estimates, express the mapping from the parameters to the optimal portfolio weights as a set of recursive functions:

$$\begin{aligned}
x^{\star}_{t+\tau-1} &= x(1, \phi, z_{t+\tau-1}, \theta) \\
x^{\star}_{t+\tau-2} &= x\left(2, \phi, z_{t+\tau-2}, \theta, x^{\star}_{t+\tau-1}\right) \\
x^{\star}_{t+\tau-3} &= x\left(3, \phi, z_{t+\tau-3}, \theta, \left\{x^{\star}_{t+\tau-1}, x^{\star}_{t+\tau-2}\right\}\right) \\
&\dots \\
x^{\star}_{t} &= x\left(\tau, \phi, z_{t}, \theta, \left\{x^{\star}_{t+\tau-1}, \dots, x^{\star}_{t+1}\right\}\right).
\end{aligned} \tag{3.11}$$

To compute the asymptotic standard errors of the estimates $\hat{x}^{\star}_t$ we also need to account for the estimation error in the preceding portfolio estimates $\{\hat{x}^{\star}_s\}_{s=t+1}^{T-1}$. This is accomplished by including in the derivatives $x_4(\cdot)$ in Eq. (3.2), also the terms:

$$\sum_{s=t+1}^{T-1} \frac{\partial x\left(t, \phi, z_t, \theta, \left\{x^{\star}_s\right\}_{s=t+1}^{T-1}\right)}{\partial x^{\star}_s} \frac{\partial x^{\star}_s}{\partial \theta}. \tag{3.12}$$

Intuitively, the longer the investment horizon, the more imprecise are the estimates of the optimal portfolio weights, because the estimation error in the sequence of optimal portfolio weights accumulates through the recursive nature of the solution.

Bayesian Estimation There is nothing inherently frequentist about the plug-in estimation. Inferences about optimal portfolio weights can be drawn equally well from a Bayesian perspective. Starting with a posterior distribution of the parameters $p(\theta|Y_T)$, use the mapping (3.1) or (3.11) to compute the posterior distribution of the portfolio weights $p(x^{\star}_t|Y_T)$ and then draw inferences about $x^{\star}_t$ using the moments of this posterior distribution.

Consider again the mean–variance problem. Assuming normally distributed returns and uninformative priors, the posterior of μ conditional on Σ^{-1}, $p(\mu|\Sigma^{-1}, Y_T)$, is Gaussian with mean $\hat{\mu}$ and covariance matrix Σ/T. The marginal posterior of Σ^{-1}, $p(\Sigma^{-1}|Y_T)$, is a Wishard distribution with mean $\overline{\Sigma}^{-1} = (T-N)\hat{S}^{-1}$ and $T-N$ degrees of freedom.[14] It follows that the posterior of the optimal portfolio weights $x^{\star} = (1/\gamma)\Sigma^{-1}\mu$, which can be computed from $p(\mu, \Sigma^{-1}|Y_T) \equiv p(\mu|\Sigma^{-1}, Y_T)\, p(\Sigma^{-1}|Y_T)$, has a mean of $(1/\gamma)\overline{\Sigma}^{-1}\hat{\mu}$.[15] As is often the case with uninformative priors, the posterior means, which are the Bayesian estimates for quadratic loss, coincide with frequentist estimates (except for the difference in degrees of freedom).

[14] See Box and Tiao (1973) for a review of Bayesian statistics.

[15] Although the posterior of $x = (1/\gamma)\Sigma^{-1}\mu$ is not particularly tractable, its mean can be easily computed using the law of iterated expectations $\mathrm{E}[\Sigma^{-1}\mu] = \mathrm{E}\big[\mathrm{E}[\Sigma^{-1}\mu|\Sigma]\big] = \mathrm{E}\big[\Sigma^{-1}\mathrm{E}[\mu|\Sigma]\big] = \mathrm{E}[\Sigma^{-1}]\hat{\mu} = \overline{\Sigma}^{-1}\hat{\mu}$.

Economic Loss How severe is the statistical error of the plug-in estimates in an economic sense? One way to answer this question is to measure the economic loss from using the plug-in estimates as opposed to the truly optimal portfolio weights. An intuitive measure of this economic loss is the difference in certainty equivalents. In the mean–variance problem (2.8), for example, the certainty equivalent of the true portfolio weights $x^\star$ is

$$\mathrm{CE} = x^{\star\prime}\mu - \frac{\gamma}{2}x^{\star\prime}\Sigma x^\star \tag{3.13}$$

and the certainty equivalent of the plug-in estimates $\hat{x}^\star$ is

$$\hat{\mathrm{CE}} = \hat{x}^{\star\prime}\mu - \frac{\gamma}{2}\hat{x}^{\star\prime}\Sigma\hat{x}^\star. \tag{3.14}$$

The certainty equivalent loss is defined as the expected difference between the two:

$$\mathrm{CE\ loss} = \mathrm{CE} - \mathrm{E}\big[\hat{\mathrm{CE}}\big], \tag{3.15}$$

where the expectation is taken with respect to the statistical error of the plug-in estimates (the certainty equivalents already capture the return uncertainty). Cho (2007) shows that this certainty equivalent loss can be approximated by:

$$\mathrm{CE} - \mathrm{E}\big[\hat{\mathrm{CE}}\big] \simeq \frac{\gamma}{2} \times \mathrm{tr}\big[\mathrm{cov}[\hat{x}^\star]\Sigma\big]. \tag{3.16}$$

The certainty equivalent loss depends on the level of risk aversion, the covariance matrix of the plug-in estimates, and the return covariance matrix. Intuitively, the consistency of the plug-in estimator implies that on average the two portfolio policies generate the same mean return, so the first terms of the certainty equivalents cancel out. The statistical error of the plug-in estimates introduces additional uncertainty in the portfolio return, referred to as *parameter uncertainty*, which is penalized by the utility function the same way as the uncertainty inherent in the optimal portfolio returns.

For the mean–variance example with a single risky asset above:

$$\mathrm{CE\ loss} \simeq \frac{\gamma}{2} \times \mathrm{var}[\hat{x}^\star]\sigma^2. \tag{3.17}$$

Plugging in the numbers from the example, the certainty equivalent of the optimal portfolio is $\mathrm{CE} = 0.533 \times 0.06 - 2.5 \times 0.533^2 \times 0.15^2 = 1.6\%$ (the investor is indifferent between the risky portfolio returns and a certain return equal to the risk-free rate plus 1.6%) and the (asymptotic) certainty equivalent loss due to statistical error under normality is $\mathrm{CE\ loss} = 2.5 \times 0.14^2 \times 0.15^2 = 0.11\%$. Notice that, although the standard error of the plug-in portfolio weights is the magnitude as the portfolio weight itself, the certainty equivalent loss is an order of magnitude smaller. This illustrates the point made in a more general context by Cochrane (1989), that for standard preferences first-order deviations from optimal decision rules tend to have only second-order utility consequences.

Given an expression for the economic loss due to parameter uncertainty, we can search for variants of the plug-in estimator that perform better in terms of their potential economic losses. This task is taken on by Kan and Zhou (2007), who consider estimators of the form $\hat{w}^{\star} = c \times \hat{\Sigma}^{-1}\hat{\mu}$ and solve for an "optimal" constant c. Optimality here is defined as the resulting estimator being admissible, which means that no other value of c generates a smaller economic loss for some values of the true μ and Σ. Their analysis can naturally be extended to estimators that have different functional forms.

3.1.2. Finite Sample Properties

Although asymptotic results are useful to characterize the statistical uncertainty of plug-in estimates, the real issue, especially for someone considering to use plug-in estimates in real-life applications, is finite-sample performance. Unfortunately, there is a long line of research documenting the shortcomings of plug-in estimates, especially in the context of large-scale mean–variance problems (e.g., Best and Grauer, 1991; Chopra and Ziemba, 1993; Jobson and Korkie, 1980, 1981; Michaud, 1989). The general conclusions from these papers is that plug-in estimates are extremely imprecise and that, even in relatively large samples, the asymptotic approximations above are quite unreliable. Moreover, the precision of plug-in estimates deteriorates drastically with the number of assets held in the portfolio. Intuitively, this is because, as the number of assets increases, the number of unique elements of the return covariance matrix increases at a *quadratic* rate. For instance, in the realistic case of 500 assets the covariance matrix involves more than 125,000 unique elements, which means that for a post-war sample of about 700 monthly returns we have less than three degrees of freedom per parameter ($500 \times 600 = 350{,}000$ observations and 125,000 parameters). I first illustrate the poor finite-sample properties of plug-in estimates through a simulation experiment and then discuss a variety of ways of dealing with this problem in practice.

Jobson–Korkie Experiment Jobson and Korkie (1980) were among the first to document the finite-sample properties of plug-in estimates. The following simulation experiment replicates their main finding. Consider 10 industry-sorted portfolios. To address the question of how reliable plug-in estimates of mean–variance efficient portfolio weights are for a given sample size, take the historical sample moments of the portfolios to be the truth and simulate independent sets of 250 hypothetical return samples of different sample sizes from a normal distribution with the true moments. For each hypothetical sample, compute again plug-in estimates of the mean–variance frontier and then evaluate how close these estimates come to the true frontier. Figures 5.4 and 5.5 illustrate the results graphically, for the unconstrained and constrained (nonnegative weights) case,

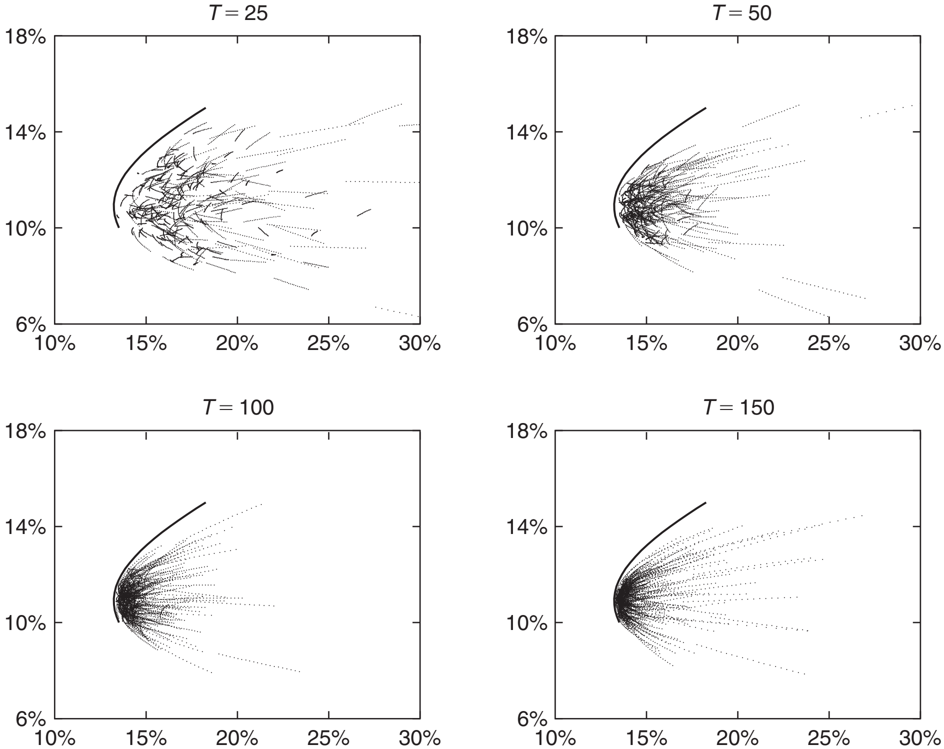

Figure 5.4 The solid line in each plot is the unconstrained mean–variance frontier for 10 industry portfolios, taking sample moments as the truth. The dotted lines show the mean–variance trade-off, evaluated using the true moments, of 250 independent plug-in estimates for 25, 50, 100, and 150 simulated returns.

respectively. Each figure shows as solid line the true mean–variance frontier and as dotted lines the mean–variance trade-off, evaluated using the true moments, of the 250 plug-in estimates for samples of 25, 50, 100, and 150 monthly returns.

The results of this experiment are striking. The mean–variance trade-off achieved by the plug-in estimates are extremely volatile and on average considerably inferior to the true mean–variance frontier. Furthermore, increasing the sample size, for example from 50 to 150, does not substantially reduce the sampling variability of the plug-in estimates. Comparing the constrained and unconstrained results, it is clear that constraints help reduce the sampling error, but clearly not to a point where one can trust the plug-in estimates, even for a sample as large as 150 months (more than 10 years of data).

To get a sense for the economic loss due to the statistical error, Fig. 5.6 shows histograms of the Sharpe ratio, again evaluated using the true moments, of the estimated unconstrained and constrained tangency portfolios for 25 and 150 observations. As a

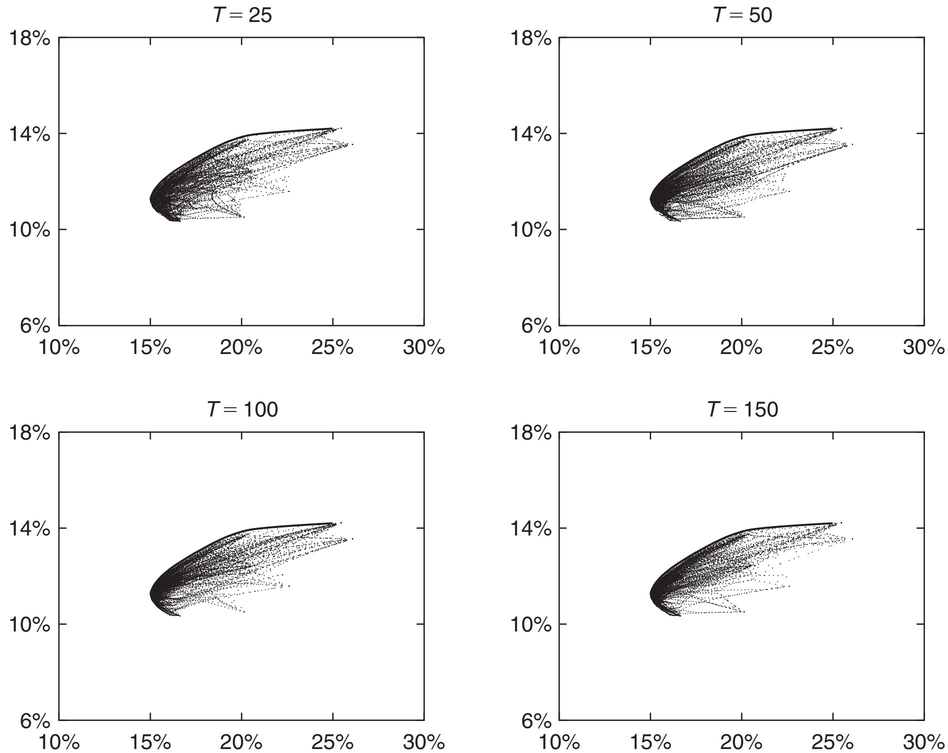

Figure 5.5 The solid line in each plot is the constrained (nonnegative portfolio weights) mean–variance frontier for 10 industry portfolios, taking sample moments as the truth. The dotted lines show the mean–variance trade-off, evaluated using the true moments, of 250 independent plug-in estimates for 25, 50, 100, and 150 simulated returns.

reference, the figure also shows as vertical lines the Sharpe ratios of the true tangency portfolio (0.61 and 0.52 for the unconstrained and constrained problems, respectively). The results in this figure are as dramatic as in the previous two figures. The Sharpe ratios of the plug-in estimates are very volatile and on average considerably lower than the truth. For example, even with 150 observations, the unconstrained Sharpe ratios have an average of 0.42 with 25th and 75th percentiles of 0.37 and 0.48, respectively. In stark contrast to the asymptotic results discussed earlier, the economic loss due to statistical error in finite samples is substantial.

In addition to being very imprecise, plug-in estimates tend to exhibit extreme portfolio weights, which, at least superficially, contradicts the notion diversification (more on this point below). For example, in the unconstrained case, the plug-in estimate of the tangency portfolio based on the historical sample moments allocates 82% to the nondurables industry and −48% to the manufacturing industry. Furthermore, the extreme portfolio

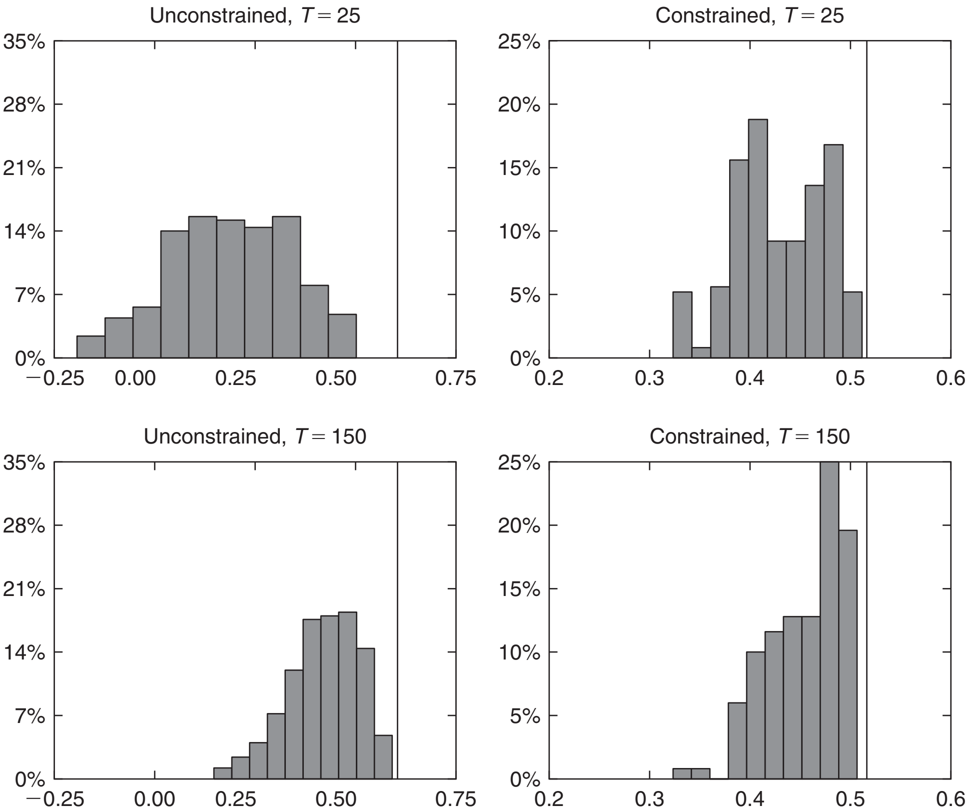

Figure 5.6 The vertical line in each plot represents the Sharpe ratio of the true unconstrained or constrained (nonnegative portfolio weights) tangency portfolios for 10 industry portfolios, taking sample moments as the truth. The histograms correspond to the Sharpe ratios, evaluated using the true moments, of 250 independent plug-in estimates for 25 or 150 simulated returns.

weights tend to be relatively unstable. Small changes in the inputs (the risk premia and covariance matrix) result in large changes in the plug-in estimates. Both of these issues have significant practical implications. Extreme positions are difficult to implement and instability causes unwarranted turnover, tax liabilities, and transaction costs. Michaud (1989) argues that extreme and unstable portfolio weights are inherent to mean–variance optimizers because they tend to assign large positive (negative) weights to securities with large positive (negative) estimation errors in the risk premium and/or large negative (positive) estimation errors in the volatility. Mean–variance optimizers therefore act as statistical "error maximizers."

Motivated by the poor finite-sample property of plug-in estimates, there exists by now an extensive literature suggesting different, but to some extent complementary, ways of improving on plug-in estimates for practical applications. These approaches include (i) shrinkage estimation, (ii) the use of factor models, and (iii) imposing portfolio constraints. I discuss each of these approaches in turn.

Shrinkage Estimation The idea of shrinkage estimation is attributed to James and Stein (1961), who noted that for $N \geq 3$ independent normal random variables, the vector of sample means $\overline{\mu}$ is dominated in terms of joint mean-squared error by a convex combination of the sample means and a common constant μ_0 (see also Efron and Morris, 1977), resulting in the estimator:

$$\mu_s = \delta\mu_0 + (1 - \delta)\overline{\mu}, \tag{3.18}$$

for $0 < \delta < 1$. The James–Stein estimator "shrinks" the sample means toward a common value, which is often chosen to be the grand mean across all variables. The estimator thereby reduces the extreme estimation errors that may occur in the cross-section of individual means, resulting in a lower overall variance of the estimators that more than compensates for the introduction of small biases. The optimal trade-off between bias and variance is achieved by an optimal shrinkage factor $\delta^\star$, given for mean-squared error loss by:

$$\delta^\star = \min\left[1, \frac{(N-2)/T}{(\overline{\mu} - \mu_0)'\Sigma^{-1}(\overline{\mu} - \mu_0)}\right]. \tag{3.19}$$

Intuitively, the optimal shrinkage factor increases in the number of means N, decreases in the sample size T (which determines the precision of the sample means), and decreases in the dispersion of the sample means $\overline{\mu}$ from the shrinkage target μ_0.

Shrinkage estimation for risk premia has been applied to portfolio choice problems by Jobson et al. (1979), Jobson and Korkie (1981), Frost and Savarino (1986), and Jorion (1986), among others. Jorion shows theoretically and in a simulation study that the optimality of the shrinkage estimator in the mean-squared error loss context considered by James and Stein (1961) carries over to estimating risk premia in the portfolio choice context. Plug-in portfolio weight estimates constructed with shrunk sample means dominate, in terms of expected utility, plug-in estimates constructed with the usual sample means.

To illustrate the potential benefits of shrinkage estimation, consider again the mean–variance example with 10 industry portfolios. Table 5.2 reports the average Sharpe ratios, evaluated using the true moments, of the 250 plug-in estimates of the unconstrained tangency portfolios for different sample sizes with and without shrinkage. To isolate the effect of statistical error in sample means, the table shows results for both a known and unknown covariance matrix. The improvement from using shrinkage is considerable. For example, with 50 observations, the average Sharpe ratio without shrinkage is 0.24

Table 5.2 Average Sharpe ratios, evaluated using the true moments, of plug-in estimates with and without shrinkage of the unconstrained tangency portfolio for 10 industry portfolios with known and unknown covariance matrix

		Known Σ		Unknown Σ	
T	Truth	Sample means	Shrinkage	Sample means	Shrinkage
25	0.624	0.190	0.428	0.169	0.270
50	0.624	0.236	0.446	0.223	0.371
100	0.624	0.313	0.477	0.298	0.443
150	0.624	0.362	0.495	0.348	0.473
250	0.624	0.418	0.512	0.411	0.501

The results are based on 250 simulated samples of size T.

or 0.22, depending on whether the covariance matrix is know or unknown, compared to the Sharpe ratio of the true tangency portfolio of 0.62. With shrinkage, in contrast, the average Sharpe ratio is 0.45 with known covariance matrix (87% improvement) and 0.37 with unknown covariance matrix (63% improvement). The average shrinkage factor with a known covariance matrix ranges from 0.78 for $T = 25$–0.71 for $T = 250$. This means that the individual sample means are shrunk about two-thirds toward a common mean across all portfolios. The reason for why shrinkage estimation is in relative terms less effective with an unknown covariance matrix is that the optimal shrinkage factor in Eq. (3.19) is evaluated with a noisy estimate of the covariance matrix, which, due to the nonlinearity of optimal shrinkage factor, results in a less shrinkage overall. In particular, the average shrinkage factor with an unknown covariance matrix is 0.51 for $T = 25$, 0.72 for $T = 100$, and 0.69 for $T = 250$.

Shrinkage estimation can also be applied to covariance matrices. In the portfolio choice context, Frost and Savarino (1986) and Ledoit and Wolf (2003, 2004) propose return covariance matrix estimators that are convex combinations of the usual sample covariance matrix $\hat{\Sigma}$ and a shrinkage target S (or its estimate $\hat{S}$):

$$\hat{\Sigma}_s = \delta\hat{S} + (1-\delta)\hat{\Sigma}. \tag{3.20}$$

Sensible shrinkage targets include an identity matrix, the covariance matrix corresponding to a single- or multifactor model, or a covariance matrix with equal correlations.

Ledoit and Wolf (2003) derive the following approximate expression for the optimal shrinkage factor assuming mean-squared error loss:

$$\delta^{\star} \simeq \frac{1}{T}\frac{A-B}{C}, \tag{3.21}$$

with

$$
\begin{aligned}
A &= \sum_{i=1}^{N}\sum_{j=1}^{N} \text{asy var}\left[\sqrt{T}\hat{\sigma}_{i,j}\right] \\
B &= \sum_{i=1}^{N}\sum_{j=1}^{N} \text{asy cov}\left[\sqrt{T}\hat{\sigma}_{i,j}, \sqrt{T}\hat{s}_{i,j}\right] \\
C &= \sum_{i=1}^{N}\sum_{j=1}^{N} \left(\sigma_{i,j} - s_{i,j}\right)^2 .
\end{aligned} \tag{3.22}
$$

The optimal shrinkage factor reflects the usual bias versus variance trade-off. It decreases in the sample size T, increases in the imprecision of $\hat{\Sigma}$ (through A), decreases in the covariance of the errors in estimates of $\hat{\Sigma}$ and $\hat{S}$ (through B), and decreases in the bias of S (through C). Ledoit and Wolf (2003) also describe how to consistently estimate the asymptotic second moments needed to evaluate the optimal shrinkage factor in practice. Finally, they show that, besides reducing sampling error, shrinkage to a positive definite target guarantees that the resulting estimate is also positive definite, even when the sample covariance matrix itself is singular (when $N > T$). This makes shrinkage a particularly practical statistical tool for constructing large-scale equity portfolios.

The idea of shrinkage estimation can in principle also be applied directly to the plug-in estimates of the optimal portfolio weights, resulting in an estimator of the form:

$$\hat{w}_s^\star = \delta w_0 + (1 - \delta)\hat{w}^\star, \tag{3.23}$$

for some sensible shrinkage target w_0. There are several potential advantages of shrinking the plug-in estimates, compared to shrinking their inputs. First, it may be easier to specify ex-ante sensible shrinkage targets, such as equal weights $1/N$ or observed relative market capitalization weights in a benchmark portfolio. Second, shrinking the plug-in estimates may be more effective because it explicitly links first and second moments. It is possible, for example, to shrink both first and second moments toward zero, thinking that the statistical error has been reduced, but leave the plug-in portfolio weights unchanged. Third, shrinkage of the plug-in estimates can be more naturally combined with an economic loss function. Specifically, the optimal shrinkage factor could be chosen to maximize the expected utility from using the shrunk plug-in estimates, as opposed to minimize its mean-squared error. Whether any of these advantages are materialized in practice remains to be seen.

Any form of shrinkage estimation involves seemingly ad-hoc choices of the shrinkage target and the degree of shrinkage (or equivalently the loss function which determines the optimal degree of shrinkage). Both of these issues are naturally resolved in a Bayesian framework, where the location of the prior beliefs can be interpreted as the shrinkage target and the variability of the prior beliefs relative to the information contained in the data automatically determines how much the estimates are shrunk toward the prior. I will return to the Bayesian interpretation of shrinkage and the choice of priors in Section 3.2.

Factor Models The second approach to reducing the statistical error of the plug-in estimates is to impose a factor structure for the covariation among assets to reduce the number of free parameters of the covariance matrix. Sharpe (1963) first proposed using the covariance matrix implied by a single-factor market model in the mean–variance problem:

$$r_{i,t} = \alpha_i + \beta_i r_{m,t} + \varepsilon_{i,t}, \tag{3.24}$$

where the residuals $\varepsilon_{i,t}$ are assumed to be uncorrelated across assets. Stacking the N market betas β_i into a vector β, the covariance matrix implied by this single-factor model is

$$\Sigma = \sigma_m^2 \beta\beta' + \Sigma_\epsilon, \tag{3.25}$$

where Σ_ϵ is a diagonal residual covariance matrix with non-zero elements $\sigma_{\varepsilon,i}^2 = \text{var}[\varepsilon_{i,t}]$. The advantage of this approach is that it reduces the dimensionality of the portfolio problem to $3N + 1$ terms $\left(\left\{\alpha_i, \beta_i, \sigma_{\varepsilon,i}^2\right\}_{i=1}^N \text{ and } \sigma_m^2\right)$. The drawback, in exchange, is that a single factor may not capture all of the covariation among assets, leading not only to a biased but potentially systematically biased estimate of the return covariance matrix.

The obvious way to overcome this drawback is to increase the number of factors capturing the covariation among assets. In a more general K-factor model:

$$r_{i,t} = \alpha_i + \beta_i' f_t + \varepsilon_{i,t}, \tag{3.26}$$

where β_i is now a vector of factor loadings, f_t is a vector a factor realizations (which still need to be specified), and the residuals $\varepsilon_{i,t}$ are again assumed to be uncorrelated across asset. The implied return covariance matrix is

$$\Sigma = B\Sigma_f B' + \Sigma_\varepsilon, \tag{3.27}$$

where B denotes the $N \times K$ matrix of stacked factor loadings, Σ_f is the covariance matrix of the factors, and Σ_ε is a diagonal residual covariance matrix. If the factors are correlated, the portfolio problem is reduced to $K(K+1)/2 + N(K+2)$ terms. If the factors are uncorrelated, which is a common assumption implying that Σ_f is also diagonal, the problem is further reduced to $K + N(K+2)$ terms. To illustrate the degree of dimension reduction achieved by multifactor models, consider again the case of 500 assets. With five factors, there are 3515 coefficients to estimate if the factors are correlated, as opposed to 125,000 in the case without factors. This translates into a more than 33-fold increase in the degrees of freedom (from less than 3 to more than 99).

The practical difficulty with implementing a multifactor model is the choice of common factors. There are essentially three ways to approach this problem. First, the choice of factors can be based on economic theory. Examples include using the market or aggregate wealth portfolio, as implied by the CAPM, which results in the approach of Sharpe (1963), or using multiple intertemporal hedge portfolios that are maximally correlated with changes in the aggregate investment opportunity set, as implied by Merton's (1973) ICAPM. Second, the choice of factors can be based on empirical work, including, for example, macroeconomic factors (e.g., Chen et al., 1986), industry factors, firm characteristic-based factors (e.g., Fama and French, 1993), and combinations thereof (e.g., BARRA's equity risk models). Third, the factors can be extracted directly from returns using a statistical procedure such as factor analysis or principal components analysis (e.g., Connor and Korajczyk, 1988). Moving from theoretical factors, to empirical factors, to statistical factors, we capture, by construction, increasingly more of the covariation among assets. In exchange, the factors become more difficult to interpret, which raises concerns about data mining.

Chan et al. (1999) study the performance of different factor model specifications in a realistic rolling-sample portfolio choice problem. Their results show that factor models clearly improve the performance of the plug-in estimates. However, no clear favorite specification emerges, both in terms of the number and the choice of factors. A simple CAPM-based single-factor model performs only marginally worse than a high-dimensional model with industry and characteristic-based factors.

Portfolio Constraints The third approach to reducing the statistical error inherent in plug-in estimation is to impose constraints on the portfolio weights. It is clear from comparing the results in Figs. 5.4 and 5.5 that imposing portfolio constraints helps. Frost and Savarino (1988) confirm this impression more scientifically by demonstrating that portfolio constraints truncate the extreme portfolio weights and thereby improve the performance of the estimates. Their results suggest that, consistent with Michaud's (1989) view of optimizers as error maximizers, the extreme portfolio weights that being truncated are associated with estimation error.

There are numerous ways of constraining portfolio weights. The most popular constraints considered in the academic literature are constraints that limit short-selling and constraints that limit the amount of borrowing to invest in risky assets. Although these constraints are obviously also very relevant in practice, realistic investment problems are subject to a host of other constraints, such as constraints on the maximum position in a single security, on the maximum exposure to a given industry or economic sector, on the liquidity of a security, or on the risk characteristics of a security. In addition, it is common practice to perform an initial screening of the universe of all securities to obtain a smaller and more manageable set of securities. These initial screens can be based on firm characteristics, including accounting and risk measures, liquidity measures, transaction cost measures, or even return forecasts.

Although portfolio constraints are an integral part of the investment process in practice, Green and Hollifield (1992) argue that, from a theoretical perspective, extreme portfolio weights do not necessarily imply that a portfolio is undiversified. The intuition of their argument is as follows. Suppose returns are generated by a single-factor model and therefore contain both of systematic and idiosyncratic risk. The aim is to minimize both sources of risk through diversification. Instead of using a mean–variance optimizer, consider an equivalent but more transparent two-step procedure in which we first diversify away idiosyncratic risk and then diversify away systematic risk. In the first step, sort stocks based on their factor loading and form equal-weighted portfolios with high factor loadings and with low factor loadings. With a large number of stocks, each of these portfolios will be well diversified and therefore only exposed to systematic risk. In the second step, take partially offsetting positions in the systematic risk portfolios to eliminate, as much as possible given the adding-up constraint on the overall portfolio weights, the systematic risk exposure. Although the outcome is a portfolio that is well diversified in terms of both idiosyncratic and systematic risk, Green and Hollifield show that the second step can involve extreme long-short positions. The implication of this argument is that, contrary to popular belief and common practice, portfolio constraints may actually hurt the performance of plug-in estimates.

Relating Shrinkage Estimation, Factor Models, and Portfolio Constraints The argument of Green and Hollifield (1992) creates tension between economic theory and the empirical fact that imposing portfolio constraints indeed improves the performance of plug-in estimates in practice. This tension is resolved by Jagannathan and Ma (2003), who show that certain constraints on the portfolio weights can be interpreted as a form of shrinkage estimation. Because shrinkage improves the finite-sample properties of plug-in estimates, it is no longer puzzling that constraints also help, even if they are not theoretically justified. As with all forms of shrinkage estimation, constrained plug-in estimates are somewhat biased but much less variable than unconstrained plug-in estimates.

Specifically, for the problem of finding a global minimum variance portfolio (in Fig. 5.1) subject to short-sale constraints $x_t \geq 0$ and position limits $x_t \leq \overline{x}$, the constrained portfolio weights x_t^+ are mathematically equivalent to the unconstrained portfolio weights for the adjusted covariance matrix:

$$\tilde{\Sigma} = \Sigma + (\delta\iota' + \iota\delta') - (\lambda\iota' + \iota'\lambda), \tag{3.28}$$

where λ is the vector a Lagrange multipliers for the short-sale constraints and δ is the vector of Lagrange multipliers for the position limits. Each Lagrange multiplier takes on a positive value whenever the corresponding constraint is binding and is equal to zero otherwise. To understand better how Eq. (3.28) amounts to shrinkage, suppose the position limit constraints are not binding but the short-sale constraint is binding for stock i, so that $\delta = 0$ and $\lambda_i > 0$. The variance of stock i is reduced to $\tilde{\sigma}_{i,i} = \sigma_{i,i} - 2\lambda_i$ and all covariance are reduced to $\tilde{\sigma}_{i,j} = \sigma_{i,j} - \lambda_i - \lambda_j$. As stocks with negative weights in minimum variance portfolios tend to have large positive covariances with other stocks, short-sale constraints effectively shrink these positive covariances toward zero. Analogously, suppose the short-sale constraints are not binding but the position limit constraint is binding for stock i, so that $\lambda = 0$ and $\delta_i > 0$. In that case, the variance of stock i is increased to $\tilde{\sigma}_{i,i} = \sigma_{i,i} + 2\delta_i$ and the covariances are all increased to $\tilde{\sigma}_{i,j} = \sigma_{i,j} + \delta_i + \delta_j$. Since stocks with large positive weights in minimum variance portfolios tend to have large negative covariances with other stocks, position limit constraints effectively shrink these negative covariances toward zero.

A similar result holds for the constrained mean–variance problem. The constrained mean–variance efficient portfolio weights x_t^+ are mathematically equivalent to the unconstrained portfolio weights for the adjusted mean vector:

$$\tilde{\mu} = \mu + \frac{1}{\lambda_0}\lambda - \frac{1}{\lambda_0}\delta \tag{3.29}$$

and adjusted target return:

$$\tilde{\overline{\mu}} = \overline{\mu} + \frac{1}{\lambda_0}\delta'\overline{x}, \tag{3.30}$$

where $\lambda_0 > 0$ is the Lagrange multiplier for the expected return constraint $x_t'\mu = \overline{\mu}$, which is always binding. If the position limit constraints are not binding but the short-sale constraint is binding for stock i, the expected return on stock i is increased to $\tilde{\mu}_i = \mu_i + \lambda_i/\lambda_0$. Since stocks with negative weights in mean–variance efficient portfolios tend to have negative expected returns, the short-sale constraints shrink the expected return toward zero. Analogously, if the short-sale constrains are not binding but the position limit constraint is binding for stock i, the expected return on stock i is decreased to $\tilde{\mu}_i = \mu_i - \delta_i/\lambda_0$. Since stocks with large positive weights tend to have large positive expected returns, position limit constraints also shrink the expected return toward zero.

3.2. Decision Theory

In the second traditional econometric approach, decision theory, the econometrician takes on the role of the investor by choosing portfolio weights that are optimal with respect to his or her subjective belief about the true return distribution.[16]

[16] An alternative way of dealing with parameter uncertainty is "robust control," where instead of improving on the statistical side of the problem, the decision maker adjusts the optimization problem. For example, in the max-min approach pioneered by Hansen and Sargent (1995), the decision maker maximizes expected utility evaluated under a worst-case return distribution (for a set of candidate distribution). See Maenhout (2004, 2006) for applications of robust control to portfolio choice problems.

In the presence of statistical uncertainty about the parameters or even about the parameterization of the data generating process, this subjective return distribution may be quite different from the results of plugging point estimates in the data generating process. As a result, the econometrician's optimal portfolio weights can also be quite different from the plug-in estimates described earlier.

3.2.1. Parameter Uncertainty

Consider, for illustrative purposes, a single-period or myopic portfolio choice with i.i.d. returns. We can write the expected utility maximization more explicitly as:

$$\max_{x_t} \int u\big(x_t' r_{t+1} + R^f\big) p(r_{t+1}|\theta) \mathrm{d}r_{t+1}, \tag{3.31}$$

where $p(r_{t+1}|\theta)$ denotes the true return distribution parameterized by θ. Until now, it was implicitly assumed that this problem is well posed, in the sense that the investor has all information required to solve it. However, suppose instead that the investor knows the parametric form of the return distribution but not the true parameter values, which, of course, is far more realistic. In that case, the problem cannot be solved as it is because the investor does not know for which parameter values θ to maximize the expected utility.

There are at least three ways for the investor to proceed. First, the investor can naively use estimates of the parameters in place of the true parameter values, analogous to the plug-in estimation approach (except now it is the investor who needs to make a decision, not an econometrician drawing inferences, relying on point estimates). The resulting portfolio weights are optimal only if the estimates happen to coincide with the true values, a zero-probability event in finite samples, and suboptimal otherwise. Second, the investor can consider the parameter values that correspond to the *worst case* outcome under some prespecified set of possible parameter values, leading to extremely conservative portfolio weights that are robust, as opposed to optimal, with respect to the uncertainty about the parameters (a decision theoretic approach called robust control). Third, the investor can eliminate the dependence of the optimization problem on the unknown parameters by replacing the true return distribution with a subjective distribution that depends only on the data the investor observes and on personal ex-ante beliefs the investor may have had about the unknown parameters before examining the data. The resulting portfolio weights are optimal with respect to this subjective return distribution but suboptimal with respect to the true return distribution. However, this suboptimality is irrelevant, in some sense, because the truth is never revealed anyway. To the extent that the subjective return distribution incorporates all of the available information (as oppose to just a point estimate or worst case outcome), this third approach is the most appealing to many.

Zellner and Chetty (1965), Klein and Bawa (1976), and Brown (1978) were among the first to advocate using subjective return distributions in portfolio choice problems. Given the data Y_T and a prior belief about of the parameters $p_0(\theta)$, the posterior distribution of the parameters is given by Bayes' theorem as:

$$p(\theta|Y_T) = \frac{p(Y_T|\theta)\, p_0(\theta)}{p(Y_T)} \propto p(Y_T|\theta)\, p_0(\theta), \tag{3.32}$$

where the distribution of the data conditional on the parameters can also be interpreted as the likelihood function $\mathcal{L}(\theta|Y_T)$. This posterior distribution can then be used to *integrate out* the unknown parameters from the return distribution to obtain the investor's *subjective* (since it involves subjective priors) return distribution:

$$p\big(r_{t+1}|Y_T\big) = \int p\big(r_{t+1}|\theta\big)p\big(\theta|Y_T\big)\mathrm{d}\theta. \tag{3.33}$$

Finally, we simply replace the true return distribution in the expected utility maximization with this subjective return distribution and solve for the optimal portfolio weights.

Formally, the investor solves the problem:

$$\max_{x_t} \int u\big(x_t' r_{t+1} + R^f\big)p\big(r_{t+1}|Y_T\big)\mathrm{d}r_{t+1}, \tag{3.34}$$

which can we can rewrite, using the construction of the posterior, as:

$$\max_{x_t} \int \left[\int u\big(x_t' r_{t+1} + R^f\big)p\big(r_{t+1}|\theta\big)\mathrm{d}r_{t+1} \right] p\big(\theta|Y_T\big)\mathrm{d}\theta. \tag{3.35}$$

Comparing Eqs. (3.31) and (3.35), it is now clear how the investor overcomes the issue of not knowing the true parameter values. Rather than solving the optimization problem for a single choice of parameter values, the investor effectively solves an *average* problem over all possible set of parameter values, where the expected utility of any given set of parameter values, the expression in brackets above, is weighted by the investor's subjective probability of these parameter values corresponding to the truth.

Uninformative Priors The choice of prior is critical in this Bayesian approach. Priors are either informative or uninformative. Uninformative priors contain little if any information about the parameters and lead to results that are comparable, but not identical in finite samples, to plug-in estimates. Consider the simplest possible example of a single i.i.d. normal return with constant mean μ and volatility σ. Assume initially that the volatility is known. Given a standard uninformative prior for the mean, $p(\mu) \propto c$, the posterior distribution of the mean is

$$p(\mu|\sigma, Y_T) = \mathrm{N}\big[\hat{\mu}, \sigma^2/T\big], \tag{3.36}$$

where $\hat{\mu}$ is the usual sample mean. This posterior distribution of the mean then implies the following predictive return distribution:

$$p(r_{T+1}|\sigma, Y_T) = \int p(r_{T+1}|\mu, \sigma)p(\mu|\sigma, Y_T)\mathrm{d}\mu = \mathrm{N}\big[\hat{\mu}, \sigma^2 + \sigma^2/T\big]. \tag{3.37}$$

Comparing this predictive return distribution to the plug-in estimate $\mathrm{N}[\hat{\mu}, \sigma^2]$ illustrates one of the effects of parameter uncertainty. In the Bayesian portfolio choice problem, the variance of returns is inflated because, intuitively, returns differ from the sample mean for two reasons. Returns have a known variance around the unknown true mean of σ^2, and the sample mean is a noisy estimate of the true mean with a variance of σ^2/T. The posterior variance of returns is therefore $\sigma^2 + \sigma^2/T$.

Relaxing the assumption of a known volatility, an uninformative prior of the form $p(\mu, \ln\sigma) = c$ leads to the joint posterior distribution of the parameters:

$$p(\mu, \sigma|Y_T) \propto \frac{1}{\sigma^{N+1}} \exp\left\{-\frac{N(\mu - \hat{\mu})^2}{2\sigma^2} - \frac{(N-1)\hat{\sigma}^2}{2\sigma^2}\right\}, \tag{3.38}$$

which, in turn, implies the following predictive return distribution:

$$p(r_{T+1}|Y_T) = \iint p(r_{T+1}|\mu, \sigma)p(\mu, \sigma|Y_T)\mathrm{d}\mu\,\mathrm{d}\sigma = t\big[\hat{\mu}, \hat{\sigma}^2 + \hat{\sigma}^2/T, N-1\big], \tag{3.39}$$

where $t[m, s^2, v]$ denotes a Student-t distribution with mean m, variance s^2, and v degrees of freedom. The mean of the predictive distribution is again the sample mean and the variance is analogous to the case with a known volatility, except with sample estimates. The only difference between the posteriors (3.37) and (3.39) is the distributional form. Specifically, since the t distribution has fatter tails than the normal distribution, especially for small degrees of freedom, parameter uncertainty about the volatility causes the tails of the posterior return distribution to fatten, relative to the case with a known volatility. Intuitively, the predictive return distribution turns into a mixture of normal distributions, each with a different volatility, as the uncertainty about the volatility is averaged out.

Although the aforementioned discussion is fairly simplistic, in that it only deals with a single risky asset and i.i.d. returns, the basic intuition extends directly to cases with multiple assets and with more complicated return models. In general, uncertainty about unconditional and/or conditional first moments tends to increase the posterior variance of returns, and uncertainty about unconditional and/or conditional second moments tends to fatten the tails of the predictive return distribution.

Equations (3.37) and (3.39) illustrate that there are differences between the Bayesian portfolio choice and plug-in estimates. However, it is important to acknowledge that, at least in this simple i.i.d. example, these differences are in practice a small-sample phenomenon. For example, suppose the volatility is known to be 18%. With only 12 observations, the posterior volatility of returns in Eq. (3.37) is equal to $\sqrt{(1+1/12)} \times 18\% = 18.75\%$. Parameter uncertainty increase the return volatility by 4%. With a more realistic sample size of 120 observations, however, the posterior volatility of returns is $\sqrt{1+1/120} \times 18\% = 18.07\%$, an increase of a negligible 0.4%. Similarly, in the case with an unknown volatility. The 5% critical value of the t distribution with 11 degrees of freedom (for $T = 12$) equals 2.18, considerably larger than 1.96 under normality. However, with 119 degrees of freedom, the critical value is 1.97, which means that the predictive distribution is virtually Gaussian (and in fact identical to its plug-in counterpart).

Guided by the long-held belief that returns unpredictable, the initial papers on parameter uncertainty were formulated in the context of i.i.d. normal returns. Following the relatively recent evidence of return predictability, Kandel and Stambaugh (1996) and Barberis (2000) reexamine the role of parameter uncertainty when returns are predictable by the dividend yield in the context of the VAR model (2.16). In particular, Barberis (2000) documents that, even in moderate size sample, parameter uncertainty can lead to substantial differences in the optimal allocation to stocks in a long-horizon portfolio choice problem. The intuition for this result is the following. As the horizon increases, the variance of returns around the true conditional mean increases linearly, because returns are conditionally uncorrelated. The variance of the estimated conditional mean around the true conditional mean, however, increases more than linearly, because the estimation error is the same in every future time period (ignoring the important issue of learning). As a result, the contribution of parameter uncertainty to the posterior variance of returns increases in relative terms as the return horizon increases.

Informative Priors Most applications of Bayesian statistics in finance employ uninformative priors, with the reasoning that empirical results with uninformative priors are most comparable to results obtained through classical statistics and therefore are easier to relate to the literature. In the context of an investor's portfolio choice problem, however, the main advantage of the Bayesian approach is the ability to incorporate subjective information through informative priors. Because portfolio choice problems are by nature subjective decision problems, not objective inference problems, there is no need to facilitate comparison.

The difficulty with using informative priors lies in maintaining analytic tractability of the posterior distributions. For this reason, the literature deals almost exclusively with so-called conjugate priors, for which the conditional posteriors are members of the same distributional class as the priors. For example, the most common conjugate prior problem involves a Gaussian likelihood function, a Gaussian prior for first moments, and an inverse gamma (or inverse Wishard in the multivariate case) prior for second moments. With this particular combination, the conditional posteriors of the first and second moments are once again Gaussian and inverse gamma, respectively. Conjugate priors are particularly convenient in problems that involve updating of previously formed posteriors with new data. In such problems, the old posterior becomes the new prior, which is then combined with the likelihood function evaluated at the new data. With conjugate priors, the updated posterior has the same distributional form as the old posterior.

To illustrate the role of informative priors and the similarities to classical shrinkage estimation, consider again the case of a single risky asset with i.i.d. normal returns and a known volatility. Assume that the investor has a normally distributed prior belief about μ centered at a prior mean of $\overline{\mu}$ with a variance of τ^2:

$$p(\mu) = \mathrm{N}[\overline{\mu}, \tau^2]. \tag{3.40}$$

Because of the conjugate structure, combining this prior with the likelihood function yields the posterior distribution:

$$p(\mu|\sigma, Y_T) = \mathrm{N}\left[\frac{\tau^2}{\tau^2 + \sigma^2/T}\hat{\mu} + \frac{\sigma^2/T}{\tau^2 + \sigma^2/T}\overline{\mu}, \frac{(\sigma^2/T)\tau^2}{\sigma^2/T + \tau^2}\right]. \tag{3.41}$$

The posterior mean is simply a relative precision weighted average of the sample and prior means. The smaller the prior uncertainty τ, the more weight is placed on the prior mean $\overline{\mu}$ and, conversely, the larger T or the smaller σ, both of which imply that the sample mean is more precisely estimated, the more weight is placed on the sample mean $\hat{\mu}$. Intuitively, the posterior mean shrinks the sample mean toward the prior mean, with the shrinkage factor depending on the relative precisions of the sample and prior means. The posterior variance is lower than the variance of the sample mean by a factor of $\tau^2/(\sigma^2/T + \tau^2)$, reflecting the fact that information is added through the informative prior. Finally, given the posterior of the mean, the predictive return distribution is obtained analogous to Eq. (3.37):

$$\begin{aligned} p(r_{T+1}|\sigma, Y_T) &= \int p(r_{T+1}|\mu, \sigma)p(\mu|\sigma, Y_T)\mathrm{d}\mu \\ &= \mathrm{N}\Bigg[\underbrace{\frac{\tau^2}{\tau^2 + \sigma^2/T}\hat{\mu} + \frac{\sigma^2/T}{\tau^2 + \sigma^2/T}\overline{\mu}}_{\mathrm{E}[\mu|\sigma, \mathcal{Y}_T]},\ \sigma^2 + \underbrace{\frac{(\sigma^2/T)\tau^2}{\sigma^2/T + \tau^2}}_{\mathrm{var}[\mu|\sigma, \mathcal{Y}_T]}\Bigg]. \end{aligned} \tag{3.42}$$

There are many ways of coming up with a subjective guess for the prior mean $\overline{\mu}$. One approach considered in the statistics literature is to take a preliminary look at the data and simply estimate the prior by maximum likelihood. Frost and Savarino (1986) apply this so-called empirical Bayes approach to the mean–variance problem. Imposing a prior belief of equal means across assets and estimating this grand mean from the data, the resulting posterior mean is remarkably similar to the James–Stein shrinkage estimator.

3.2.2. Incorporating Economic Views and Models

Arguably a more intuitive and certainly a more popular way of specifying a prior in the portfolio choice context is to rely on the theoretical implications of an economic model. The most famous example of this approach is Black and Litterman (1992), who use as prior the risk premia implied by mean–variance preferences and market equilibrium. Before elaborating on their model and two other examples of incorporating economic models, I describe a more general framework for combining two sources of information about expected returns, through Bayes theorem, into a single predictive return distribution.

Mixed Estimation Mixed estimation was first developed by Theil and Goldberger (1961) as a way to update the Bayesian inferences drawn from old data with the information contained in a set of new data. It applies more generally, however, to the problem of combining information from two data sources into a single posterior distribution. The following description of mixed estimation is tailored to a return forecasting problem and follows closely the econometric framework underlying the Black–Litterman model (GSAM Quantitative Strategies Group, 2000). A very similar setup is described by Scowcroft and Sefton (2003).

Assume excess returns are i.i.d. normal:

$$r_{t+1} \sim \mathrm{MVN}[\mu, \Sigma]. \tag{3.43}$$

The investor starts with a set of benchmark beliefs about the risk premia:

$$p(\mu) = \mathrm{MVN}[\overline{\mu}, \Lambda]. \tag{3.44}$$

which can be based on theoretical predictions, previous empirical analysis, or dated forecasts. In addition to these benchmark beliefs, the investor has a set of new views or forecasts v about a subset of $K \leq N$ linear combinations of returns $P\, r_{t+1}$, where P is a $K \times N$ matrix selecting and combining returns into portfolios for which the investor is able to express views. The new views are assumed to be unbiased but imprecise, with distribution:

$$p(v|\mu) = \mathrm{MVN}[P\mu, \Omega]. \tag{3.45}$$

Besides the benchmark beliefs, the estimator requires three inputs: the portfolio selection matrix P, the portfolio return forecasts v, and the forecast error covariance matrix Ω.

To demonstrate the flexibility of this specification, suppose there are three assets. The investor somehow forecasts the risk premium of the first two assets to be 5% and 15%, but, for whatever reason, is unable or unwilling to express a view on the risk premium of the third asset. This scenario corresponds to:

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad \text{and} \quad v = \begin{bmatrix} 0.05 \\ 0.15 \end{bmatrix}. \tag{3.46}$$

If instead of expressing views on the levels of the risk premia, the investor can only forecast the difference between the risk premia to be 10%, the matrices are

$$P = \begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \quad \text{and} \quad v = \begin{bmatrix} -0.10 \end{bmatrix}. \tag{3.47}$$

Once the views have been formalized, the investor also needs to specify their accuracy and correlations through the choice of Ω. In the first scenario, for instance, the investor might be highly confident in the forecast of the first risk premium, with a 1% forecasts error volatility, but less certain about the forecast of the second risk premium, with a 10% forecast error volatility. Assuming further that the two forecasts are obtained independently, the covariance matrix of the forecast errors is

$$\Omega = \begin{bmatrix} 0.01^2 & 0 \\ 0 & 0.10^2 \end{bmatrix}. \tag{3.48}$$

The off-diagonal elements of Ω capture correlations between the forecasts. Specifically, high confidence in the forecast of $\mu_1 - \mu_2$ is intuitively equivalent to very low confidence in the forecasts of μ_1 and μ_2, but with a high correlation between the two forecast errors.

Combining Eqs (3.45) and (3.45) using Bayes' theorem:

$$\begin{aligned} p(\mu|v) &\propto p(v|\mu)\, p(\mu) \\ &= \text{MVN}\big[\text{E}[\mu|v], \text{var}[\mu, v]\big], \end{aligned} \tag{3.49}$$

where the posterior moments of μ are given by:

$$\begin{aligned} \text{E}[\mu|v] &= \big[\Lambda^{-1} + P'\Omega P\big]^{-1}\big[\Lambda^{-1}\overline{\mu} + P'\Omega^{-1}v\big] \\ \text{var}[\mu|v] &= \big[\Lambda^{-1} + P'\Omega P\big]^{-1}. \end{aligned} \tag{3.50}$$

Finally, assuming Σ is known, the predictive return distribution is given by:

$$p(r_{T+1}|v) = \text{MVN}\Big[\text{E}[\mu|v], \big[\Sigma^{-1} + \text{var}[\mu|v]^{-1}\big]^{-1}\Big]. \tag{3.51}$$

Alternatively, if Σ is unknown, the predictive return distribution with conjugate prior for the covariance matrix is multivariate t with the same first and second moments, analogous to the univariate case in Eq. (3.39).

As in the more general case of informative priors, the posterior mean is simply a relative precision weighted average of the benchmark means $\overline{\mu}$ and the forecasts v (a form of shrinkage). The advantage of this particular mixed estimation setup is the ability to input forecasts of subsets and linear combinations of the risk premia. This is particularly relevant in real-life applications where forecasting the returns on every security in the investable universe (e.g., AMEX, NASDAQ, and NYSE) is practically impossible.

Black–Litterman Model The Black and Litterman (1992) model is an application of this mixed estimation approach using economically motivated benchmark beliefs $p(\mu)$ and proprietary forecasts v (obtained through empirical studies, security analysis, or other forecasting techniques). The benchmark beliefs are obtained by inferring the risk premia that would induce a mean–variance investor to hold all assets in proportion to their observed market capitalizations. Since such risk premia clear the market by setting the supply of shares equal to demand at the current price, they are labeled equilibrium risk premia.

More specifically, the equilibrium risk premia are calculated by reversing the inputs and outputs of the mean–variance optimization problem. In the mean–variance problem (2.8), the inputs are the mean vector μ and covariance matrix Σ. The output is the vector of optimal portfolio weights $x^\star = (1/\gamma)\Sigma^{-1}\mu$. Now suppose that the market as a whole acts as a mean–variance optimizer, then, in equilibrium, the risk premia and covariance matrix must be such that the corresponding optimal portfolio weights equal the observed market capitalization weights, denoted $x^\star_{\text{mkt}}$. Assuming a known covariance matrix, the relationship between the market capitalization weights and the equilibrium risk premia μ_{equil} is therefore given by $x^\star_{\text{mkt}} = (1/\gamma)\Sigma^{-1}\mu_{\text{equil}}$. Solving for the equilibrium risk premia:

$$\mu_{\text{equil}} = \gamma\Sigma\, x^\star_{\text{mkt}}. \tag{3.52}$$

The inputs to this calculation are the market capitalization weights, return covariance matrix, and aggregate risk aversion γ. The output is a vector of implied equilibrium risk premia.

Black and Litterman (1992) center the benchmark beliefs at these equilibrium risk premia and assume a precision matrix Λ proportional to the return covariance matrix Σ:

$$p(\mu) = \text{MVN}[\mu_{\text{equil}}, \lambda\Sigma]. \tag{3.53}$$

The constant λ measures the strength of the investor's belief in equilibrium. For instance, a value of $\lambda = 1/T$ places the benchmark beliefs on equal footing with sample means. Combining the benchmark beliefs with proprietary views v results in a posterior distribution for the risk premia with the following moments:

$$\begin{aligned}
\text{E}[\mu|v] &= \left[(\lambda\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}\left[(\lambda\Sigma)^{-1}\mu_{\text{equil}} + P'\Omega^{-1}v\right] \\
&= \left[(\lambda\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1}\left[\frac{\gamma}{\lambda}x^\star_{\text{mkt}} + P'\Omega^{-1}v\right] \\
\text{var}[\mu|v] &= \left[(\lambda\Sigma)^{-1} + P'\Omega^{-1}P\right]^{-1},
\end{aligned} \tag{3.54}$$

where the second line for the posterior mean, which follows from substituting Eq. (3.52) into the first line, makes clear the dependence of the mixed estimator on the observed market capitalization weights.

The idea of implied equilibrium risk premia is best illustrated through an example. Table 5.3 presents descriptive statistics for the returns on six size and book-to-market sorted stock portfolios. Table 5.4 shows in the third column the corresponding market capitalization weights for December 2003 and in the next four columns the equilibrium risk premia implied by the covariance matrix from Table 5.3 and relative risk aversion ranging from $\gamma = 1$ to $\gamma = 7.5$. For comparison, the last column repeats the sample risk premia from Table 5.3.

The results in the second table illustrate two important features of the implied equilibrium risk premia. First, the levels of the risk premia depend on the level of risk aversion, which therefore needs to be calibrated before using the results in the mixed estimator. One way to calibrate γ is to set the implied Sharpe ratio of the market portfolio to a sensible level. For instance, with $\gamma = 5$ the annualized Sharpe ratio of the market portfolio is 0.78, which is reasonable though still on the high side of historical experience for the market index. The second striking result in the table is that the implied equilibrium

Table 5.3 Descriptive statistics of six portfolios of all AMEX, NASDAQ, and NYSE stocks sorted by their market capitalization and book-to-market ratio

Size	Book to market	Risk premia (%)	Volatility (%)	Correlations				
Small	Low	5.61	24.56	1				
Small	Medium	12.75	17.01	0.926	1			
Small	High	14.36	16.46	0.859	0.966	1		
Big	Low	9.72	17.07	0.784	0.763	0.711	1	
Big	Medium	10.59	15.05	0.643	0.768	0.763	0.847	1
Big	High	10.44	13.89	0.555	0.698	0.735	0.753	0.913

Monthly data from January 1983 through December 2003.

Table 5.4 Equilibrium risk premia implied by market capitalization weights of six portfolios of all AMEX, NASDAQ, and NYSE stocks sorted by their market capitalization and book-to-market ratio on December 2003 and mean–variance preferences with different levels of risk aversion

			Equilibrium risk premia (%)				
Size	Book to market	Market weight (%)	$\gamma = 1$	$\gamma = 2.5$	$\gamma = 5$	$\gamma = 7.5$	Historical risk premia (%)
Small	Low	2.89	3.07	7.69	15.37	23.06	5.61
Small	Medium	3.89	2.21	5.52	11.03	16.55	12.75
Small	High	2.21	2.04	5.11	10.22	15.33	14.36
Big	Low	59.07	2.62	6.55	13.10	19.64	9.72
Big	Medium	23.26	2.18	5.44	10.88	16.32	10.59
Big	High	8.60	1.97	4.91	9.83	14.74	10.44

risk premia are quite different from the empirical risk premia, in particular for the small and low book-to-market portfolio. In fact, the two sets of risk premia are *negatively* correlated in the cross-section (a correlation coefficient of -0.83). A mixed estimator that places equal weights on the equilibrium risk premia and the sample risk premia, which corresponds to using $\lambda = 1/T$ and historical moments for ν, therefore generates return forecasts that are substantially less variable in the cross-section than either the equilibrium risk premia or the sample risk premia.

Return Forecasting with a Belief in No Predictability Another interesting example of incorporating economic views is the problem of forecasting returns with an prior belief in no predictability, studied by Kandel and Stambaugh (1996) as well as Connor (1997). Consider the regression:[17]

$$r_{t+1} = a + b\, z_t + \varepsilon_{t+1}, \tag{3.55}$$

where $\varepsilon_{t+1} \sim \mathrm{N}\big[0, \sigma_\varepsilon^2\big]$ and z_t are assumed exogenous with zero mean and a variance of σ_z^2. Using a standard OLS approach, the one-period ahead return forecast is given by $\hat{a} + \hat{b}\, z_T$, with $\hat{b}_{\mathrm{ols}} = \hat{\sigma}_{z,r}\big/\hat{\sigma}_z^2$. Unfortunately, this forecast tends to be very noisy because the regression usually has an R^2 around 1% and a t-statistic of the slope coefficient close to two. The potential for large estimation error renders the forecast practically useless, particularly when the forecast is used as an input to an error maximizing portfolio optimizer.

Kandel and Stambaugh (1996) and Connor (1997) recommend imposing an informative prior centered on the case of no predictability, which implies that the slope coefficient should be zero. Specifically, using the prior $p(b) = \mathrm{N}\big[0, \sigma_b^2\big]$ in a standard Bayesian regression setup yields a posterior of the slope coefficient with a mean of:

$$\hat{b}_{\mathrm{Bayes}} = \left[\frac{T\hat{\sigma}_z^2\big/\hat{\sigma}_\varepsilon^2}{\big(T\hat{\sigma}_z^2/\hat{\sigma}_\varepsilon^2\big) + \big(1/\sigma_b^2\big)}\right]\hat{b}_{\mathrm{ols}}. \tag{3.56}$$

As expected, the OLS estimate is shrunk toward the prior mean of zero, with a shrinkage factor that depends on the relative precisions of the OLS estimate and the prior mean. The critical ingredient of this approach is obviously the prior variance σ_b^2.

Because it is difficult to specify a sensible value for this prior variance ex-ante, especially without knowing σ_r^2 and σ_z^2, Connor (1997) reformulates the problem in a more intuitive and practical way. Define:

$$\rho = \mathrm{E}\left[\frac{R^2}{1 - R^2}\right], \tag{3.57}$$

[17] Although often associated in the literature, no predictability does not necessarily corresponding to market efficiency. In particular, returns can well be predictable in an efficient market with time-varying preferences or fundamental uncertainty.

Table 5.5 Shrinkage factor for the slope coefficient of a univariate return forecast regression with belief in market efficiency for different sample sizes and expected degrees of return predictability

$\rho \simeq E[R^2]$ (%)	$T = 24$	$T = 48$	$T = 60$	$T = 120$
0.50	0.11	0.19	0.23	0.38
0.75	0.15	0.26	0.31	0.47
1.00	0.19	0.32	0.38	0.55
2.00	0.32	0.49	0.55	0.71
3.00	0.42	0.59	0.64	0.78

which, for the low values of R^2 we observe in practice, is approximately equal to the expected degree of predictability $\mathrm{E}[R^2]$. Equation (3.56) can then be rewritten as:

$$\hat{b}_{\text{Bayes}} = \left[\frac{T}{T + (1/\rho)}\right]\hat{b}_{\text{ols}}, \tag{3.58}$$

where the degree of shrinkage toward zero depends only on the sample size T and on the expected degree of predictability ρ.

The appealing feature of the alternative formulation (3.58) is that the shrinkage factor applies generically to any returns forecasting regression with a prior belief in no predictability (or a regression slope of zero). Table 5.5 evaluates the shrinkage factor for different sample sizes and expected degrees of predictability. The extend of shrinkage toward zero is striking. With a realistic expected R^2 of 1% and a sample size between 5 and 10 years, the OLS estimate is shrunk roughly half-way toward zero (62% for $T = 60$ and 45% for $T = 120$).

Connor (1997) further shows that in the case of a multivariate return forecast regression $r_{t+1} = b' z_t + \varepsilon_{t+1}$, the shrinkage factor applied to each slope coefficient is also given by Eq. (3.58), except that the expected degree of return predictability ρ is replaced by a "marginal" counterpart ρ_i. This marginal expected degree of return predictability simply measures the marginal contribution of variable i to the expected regression R^2. For example, suppose the expected R^2 of a regression with three predictors is 1% and $T = 60$. If each variable contributes equally to the overall predictability, $\rho_i = 0.33\%$ and each slope coefficient is shrunk about 84% toward zero. In contrast, if the first variable accounts for 2/3 of the overall predictability, its slope coefficient is only shrunk 71% toward zero.

Cross-Sectional Portfolio Choice with a Belief in an Asset Pricing Model The third example of incorporation economic beliefs, this time originating from an equilibrium asset pricing model, is formulated by Pastor (2000). Suppose returns are generated by a single-factor model:

$$r_{i,t+1} = \alpha_i + \beta_i\, r_{m,t+1} + \varepsilon_{i,t+1} \tag{3.59}$$

with uncorrelated residuals $\varepsilon_{i,t+1} \sim \mathrm{N}[0, \sigma_\varepsilon^2]$. The theoretical prediction of the CAPM is that differences in expected returns in the cross-section are fully captured by differences in market betas and that $\alpha_i = 0$, for all stocks i. Therefore, an investor's ex-ante belief in the CAPM can be captured through an informative prior for the stacked intercepts α:

$$p(\alpha) = \mathrm{MVN}[0, \sigma_\alpha I]. \tag{3.60}$$

This prior is centered at zero, the theoretical prediction of the CAPM, with a dispersion σ_α measuring the strength of the investor's belief in the equilibrium model.

Combining the informative prior (3.60) with uninformative priors for the market betas and residual variances, the resulting posterior distribution has the following means:

$$\begin{aligned} \mathrm{E}[\alpha | Y_T] &= (1 - \delta)\hat{\alpha}_{\mathrm{ols}} \\ \mathrm{E}[\beta | Y_T] &= \hat{\beta}_{\mathrm{ols}} + \xi \end{aligned} \tag{3.61}$$

Intuitively, the intercepts are shrunk toward zero with the shrinkage factor δ depending, as usual, on T, σ_m^2, σ_ε^2, and σ_α^2. However, the problem is somewhat more complicated because, as the intercepts are shrunk toward zero, the market betas also change by ξ to better fit the cross-sectional differences in expected returns. Pastor (2000) provides expressions for δ and ξ and also considers the case of multifactor asset pricing models. Further extensions and applications are pursued by Pastor and Stambaugh (2000, 2002) and Avramov (2004).

3.2.3. Model Uncertainty

The idea of dealing with parameter uncertainty by averaging the return distribution over plausible parameter values can be naturally extended to dealing with model uncertainty by averaging over plausible model specifications. Define a model M_j as being a particular specification of the conditional return distribution and consider a finite set of J models containing the true model $M \in \{M_1, M_2, \ldots, M_J\}$. For any model j, the return distribution is $p(r_{t+1}|M_j, \theta_j)$, where the parameter vector θ_j can have different dimensions across models. Analogous to parameter uncertainty, the problem of model uncertainty is that the investor does not know which of the models to use in the portfolio choice problem.

Assume the investor can express a prior belief about each model j being the true data generator, $p(M_j)$, as well as a prior belief about the parameters of each model, $p(\theta_j|M_j)$. Combining these priors and the likelihood function, $p(Y_T|M_j, \theta_j)$, Bayes' theorem implies for each model the following posterior model probability:

$$p(M_j|Y_T) = \frac{p(Y_T|M_j)\, p(M_j)}{\sum_{j=1}^{J} p(Y_T|M_j)\, p(M_j)}, \tag{3.62}$$

where

$$p(Y_T|M_j) = \int p(Y_T|M_j, \theta_j)\, p(\theta_j|M_k) \mathrm{d}\theta_j \tag{3.63}$$

denotes the marginal likelihood of model j after integrating out the parameters θ_j.

The posterior model probabilities serve a number of purposes. First, they help to characterize the degree of model uncertainty. For instance, suppose there are five plausible models. Model uncertainty is obviously more prevalent when each model has a posterior probability of 20%, than when one model dominates with a posterior probability of 90%. Second, the posterior model probabilities can be used to select a model with highest posterior probability, or to eliminate models with negligible probabilities from the set of all models, thereby reducing the inherently high dimensionality of model uncertainty. Third, the posterior model probabilities can be used to construct a predictive return distribution by averaging across all models according to their posterior probabilities. This so-called model averaging approach is particularly useful when the degree of model uncertainty is too high for the investor to confidently single out a model as being the true data generator. Model averaging is analogous to averaging the return distribution over all parameter values according to the posterior distribution of the parameters [as in Eq. (3.35)].

Formally, we construct the following posterior probability weighted average return distribution:

$$p(r_{T+1}|Y_T) = \sum_{j=1}^{J} p(r_{T+1}|Y_T, M_j)\, p(M_j|Y_T), \tag{3.64}$$

where

$$p(r_{T+1}|Y_T, M_j) = \int p(r_{t+1}|M_j, \theta_j)\, p(\theta_j, Y_T, M_j) \mathrm{d}\theta_j \tag{3.65}$$

denotes the marginal return distribution after integrating out the parameters θ_j. An extremely convenient property of this averaged predictive return distribution is that, due to the linearity of the average, all noncentral moments are also model-averaged:

$$\mathrm{E}\big[r_{T+1}^q \big| Y_T\big] = \sum_{j=1}^{J} \mathrm{E}\big[r_{T+1}^q \big| Y_T, M_j\big]\, p(M_j|Y_T), \tag{3.66}$$

for any order q. Equation (3.66) can be used to construct (subjective) mean–variance efficient portfolio weights using as inputs the posterior return moments implied by each model as well as the posterior model probabilities.

Although intuitive and theoretically elegant, the practical implementation of model averaging is less straightforward, both from a computational and conceptual perspective. There are at least two computational issues. First, the marginal distributions (3.63) and (3.65) are typically analytically intractable and need to be evaluated numerically. Second, even in the context of linear regression models, which are most common in practice, the model space with K regressors contains 2^K permutations, for which the marginal distributions have to be evaluated (numerically). With 15 regressors, a relatively modest number, there are over 32,000 models to consider. Both of these issues can be overcome, with some effort, using the Markov chain Monte Carlo (MCMC) approach of George and McCulloch (1993).

The conceptual difficulties lie in the choice of the model set and the choice of the model priors, which are intimately related issues. By having to specify ex-ante the list of all plausible models, the investor explicitly rules out all nonincluded models (by essentially setting the prior probabilities of those models to zero). Given the existing disagreement about return modeling in the literature, it is hard to imagine that any model can be ruled out ex-ante with certainty. As for the form of the priors, an obvious choice is an uninformative prior assigning equal probabilities to all models. However, such uniform prior may actually be surprisingly informative about certain subsets of models. For example, consider a linear forecasting regression framework with K regressors. Only one of the 2^K models does not include any forecasters and is therefore consistent with the notion of market efficiency. The remaining models all exhibit some violation of market efficiency. With equal priors of $1/2^K$ for each model, the implied prior odds against market efficiency are an overwhelming $(2^K - 1)$ to one. An economically more intuitive prior might assign a probability of 1/2 to the no-predictability case and distribute the remaining probability of 1/2 evenly across all other model. Unfortunately, even this approach does not fully resolve the issue. Suppose that two-thirds of the K predictors are (highly correlated) price-scaled variables (e.g., dividend yield, earnings yield, book-to-market) and one-third are (highly correlated) interest rate variables (e.g., short rate, long rate). In that case, an evenly distributed prior across all models with predictability assigns odds of 3:2 in favor of predictability due to price-scaled variables as opposed to interest rate variables. The point of this example is to illustrate that the choice of model priors is a tricky issue that requires careful economic reasoning.

There have been a number of recent applications of model averaging to portfolio choice. Specifically, Avramov (2002) and Cremers (2002) both consider model uncertainty in linear return forecasting models. Tu and Zhou (2004) considers uncertainty about the shape of the return distribution in cross-sectional applications, and Nigmatullin (2003) introduces model uncertainty in the nonparametric approach of Aït-Sahalia and Brandt (2001) (discussed further below). The fundamental conclusion of all of these papers is that model uncertainty contributes considerably to the subjective uncertainty faced by an investor. For example, Avramov (2002) demonstrates that the contribution of model uncertainty to the posterior variance of returns is as large or even larger than the contribution of parameter uncertainty discussed earlier. It is clear from this recent literature that model uncertainty is an important econometric aspect of portfolio choice.

4. ALTERNATIVE ECONOMETRIC APPROACH

The traditional econometric approach is fundamentally a two-step procedure. In the first step, the econometrician or investor models and draws inferences about the data generating process (either through plug-in estimation or by forming a subjective belief) to ultimately, in the second step, solve for the optimal portfolio weights. The majority of my own research on portfolio choice has focused on ways to skip the first step of modeling returns and directly draw inferences about the optimal portfolio weights from the data.

Besides the obvious fact that the optimal portfolio weights are the ultimate object of interest, there are at least three other benefits from focusing directly on the portfolio weights. First, the return modeling step is without doubt the Achilles' heel of the traditional econometric approach. There is vast disagreement even among finance academicians on how to best model returns, and the documented empirical relationships between economic state variables (forecasters) and return moments are usually quite tenuous. Combined, this leads to substantial risk of severe model mispecification and estimation error, which are subsequently accentuated by the portfolio optimizer in the second step of the procedure. The intuition underlying my research is that optimal portfolio weights are easier to model and estimate than conditional return distributions. A second but related benefit of focusing on the portfolio weights is dimension reduction. Consider once again an unconditional mean–variance problem with 500 assets. The return modeling step involves more than 125,000 parameters, but the end-result of the two-step procedure are only 500 optimal portfolio weights. Focusing directly on the optimal portfolio weights therefore reduced considerably the room for model mispecification and estimation error. Third, drawing inferences about optimal portfolio weights lends itself naturally to using an expected utility-based loss function in a classical setting, as opposed to the obviously inconsistent practice of using standard squared error loss to estimate the return model in the first step and then switching to an expected utility function to solve for the optimal portfolio weights in the second step.

4.1. Parametric Portfolio Weights

The simplest way to directly estimate optimal portfolio weights is to parameterize the portfolio weights as functions of observable quantities (economic state variables and/or firm characteristics) and then solve for the parameters that maximize expected utility. This idea is developed in the context of single and multiperiod market timing problems by Brandt and Santa-Clara (2006) and in the context of a large cross-sectional portfolio choice problem by Brandt et al. (2009). Since the implementations in these two papers are somewhat different, yet complimentary, I explain each in turn.

4.1.1. Conditional Portfolio Choice by Augmenting the Asset Space

In Brandt and Santa-Clara (2006), we solve a market timing problem with parameterized portfolio weights of the form $x_t = \theta z_t$. We demonstrate that solving a conditional problem with parameterized portfolio weights is mathematically equivalent to solving an unconditional problem with an augmented asset space that includes naively managed zero-investment portfolios with excess returns of the form z_t times the excess return of each basis asset. This makes implementing our approach to dynamic portfolio choice no more difficult than implementing the standard Markowitz problem.

Consider first a single-period mean–variance problem. Assuming that the optimal portfolio weights are linear functions of K state variables z_t (which generally include a constant):

$$x_t = \theta z_t, \tag{4.1}$$

where θ is a $N \times K$ matrix of coefficients, the investor's conditional optimization problem is

$$\max_{\theta} \mathrm{E}_t\big[(\theta z_t)' r_{t+1}\big] - \frac{\gamma}{2} \operatorname{var}_t\big[(\theta z_t)' r_{t+1}\big]. \tag{4.2}$$

We use the following result from linear algebra:

$$(\theta z_t)' r_{t+1} = z_t' \theta' r_{t+1} = \operatorname{vec}(\theta)'(z_t \otimes r_{t+1}), \tag{4.3}$$

where $\operatorname{vec}(\theta)$ stacks the columns of θ and $\otimes$ denotes a Kronecker product, and define:

$$\begin{aligned} \tilde{x} &= \operatorname{vec}(\theta) \\ \tilde{r}_{t+1} &= z_t \otimes r_{t+1}. \end{aligned} \tag{4.4}$$

The investor's conditional problem can then be written as:

$$\max_{\tilde{x}} \mathrm{E}_t\big[\tilde{x}' \tilde{r}_{t+1}\big] - \frac{\gamma}{2} \operatorname{var}_t\big[\tilde{x}' \tilde{r}_{t+1}\big]. \tag{4.5}$$

Since the same $\tilde{x}$ maximizes the conditional mean–variance tradeoff at all dates t (hence no time-subscript), it also maximizes the unconditional mean–variance tradeoff:

$$\max_{\tilde{x}} \mathrm{E}\big[\tilde{x}' \tilde{r}_{t+1}\big] - \frac{\gamma}{2} \operatorname{var}\big[\tilde{x}' \tilde{r}_{t+1}\big], \tag{4.6}$$

which corresponds simply to the problem of finding the unconditional mean–variance optimal portfolio weights $\tilde{x}$ for the expanded set of $N \times K$ assets with returns $\tilde{r}_{t+1}$. The expanded set of assets can be interpreted as managed portfolios, each of which invests in a single basis asset an amount proportional to the value of one of the state variables. We therefore label these expanded set of assets "conditional portfolios." Given the solution to the unconditional mean–variance problem:

$$\tilde{x}^{\star} = \frac{1}{\gamma} \operatorname{var}[\tilde{r}_{t+1}]^{-1} \mathrm{E}[r_{t+1}], \tag{4.7}$$

we recover the conditional weight invested in each of the basis assets at any time t by simply adding up the corresponding products of elements of $\tilde{x}^\star$ and z_t in Eq. (4.1).

The idea of augmenting the asset space with naively managed portfolios extends to the multiperiod case. For example, consider a two-period mean–variance problem:

$$\max \mathrm{E}_t\big[r_{p,t\to t+2}\big] - \frac{\gamma}{2}\,\mathrm{var}_t\big[r_{p,t\to t+2}\big], \tag{4.8}$$

where $r_{p,t\to t+2}$ denotes the excess portfolio return of a two-period investment strategy:

$$\begin{aligned} r_{p,t\to t+2} &= \big(R_t^f + x_t' r_{t+1}\big)\big(R_{t+1}^f + x_{t+1}' r_{t+2}\big) - R_t^f R_{t+1}^f \\ &= x_t'\big(R_{t+1}^f r_{t+1}\big) + x_{t+1}'\big(R_t^f r_{t+2}\big) + \big(x_t' r_{t+1}\big)\big(x_{t+1}' r_{t+2}\big). \end{aligned} \tag{4.9}$$

The first line of this equation shows that $r_{p,t\to t+2}$ is a two-period excess return. The investor borrows a dollar at date t and allocates it to the risky and risk-free assets according to the first-period portfolio weights x_t. At $t+1$, the one-dollar investment results in $\big(R_t^f + x_t^\top r_{t+1}\big)$ dollars, which the investor then allocates again to the risky and risk-free assets according to the second-period portfolio weights x_{t+1}. Finally, at $t+2$, the investor has $\big(R_t^f + x_t^\top r_{t+1}\big)\big(R_{t+1}^f + x_{t+1}^\top r_{t+2}\big)$ dollars but pays $R_t^f R_{t+1}^f$ dollars for the principal and interest of the one-dollar loan. The second line of the equation decomposes the two-period excess return into three terms. The first two terms have a natural interpretation as the excess return of investing in the risk-free rate in the first (second) period and in the risky asset in the second (first) period. The third term captures the effect of compounding. Comparing the first two terms to the third, the latter is two orders of magnitude smaller than the former. The return $\big(x_t^\top r_{t+1}\big)\big(x_{t+1}^\top r_{t+2}\big)$ is a product of two single-period excess returns, which means that its units are of the order of 1/100th of a percent per year. The returns on the first two portfolios, in contrast, are products of a gross return $\big(R_t^f$ or $R_{t+1}^f\big)$ and an excess return $\big(r_{t+1}$ or $r_{t+2}\big)$, so their units are likely to be percent per year. Given that the compounding term is orders of magnitude smaller, we suggest to ignore it.

Without the compounding term, the two-period problem involves simply a choice between two intertemporal portfolios, one that holds the risky asset in the first period only and the other that holds the risky asset in the second period only. Using these two intertemporal portfolios, which we label "timing portfolios," we can solve the dynamic problem as a static mean–variance optimization. The solution is

$$\tilde{x}^\star = \frac{1}{\gamma}\mathrm{var}[\tilde{r}_{t\to t+2}]^{-1}\,\mathrm{E}[\tilde{r}_{t\to t+2}], \tag{4.10}$$

with $\tilde{r}_{t\to t+2} = \big[R_{t+1}^f r_{t+1}, R_t^f r_{t+2}\big]$. The first N elements of $\tilde{x}$, corresponding to $R_{t+1}^f r_{t+1}$, represents the fraction of wealth invested in the risky assets in the first period, and the remaining elements, corresponding to $R_t^f r_{t+2}$, are for the risky assets in the second period.

In a general H-period problem, we proceed in exactly the same way. We construct a set of timing portfolios:

$$\tilde{r}_{t\to t+H} = \left\{ \prod_{\substack{i=0 \\ i\neq j}}^{H-1} R^f_{t+i} r_{t+j+1} \right\}_{j=0}^{H-1}, \tag{4.11}$$

where each term represents a portfolio that invests in risky assets in period $t+j$ and in the risk-free rate in all other periods $t+i$, with $i \neq j$, and obtain the mean–variance solution:

$$\tilde{x}^\star = \frac{1}{\gamma} \operatorname{var}[\tilde{r}_{t\to t+H}]^{-1} \operatorname{E}[\tilde{r}_{t\to t+H}] \tag{4.12}$$

In addition, we can naturally combine the ideas of conditional and timing portfolios. For this, we simply replace the risky returns r_{t+j+1} in Eq. (4.11) with the conditional portfolio returns $z_{t+j} \otimes r_{t+j+1}$. The resulting optimal portfolio weights then provide the optimal allocations to the conditional portfolios at each date $t+j$.

The critical property of the solutions (4.7) and (4.12) is that they depend only on the unconditional moments of the expanded set of assets and therefore do not require any assumptions about the conditional joint distribution of the returns and state variables (besides that the unconditional moments exist). In particular, the solutions do not require any assumptions about how the conditional moments of returns depend on the state variables or how the state variables evolve through time. Furthermore, the state variables can predict time-variation in the first, second, and, if we consider more general utility functions, even higher-order moments of returns. Notice also that the assumption and the optimal portfolio weights are linear functions of the state variables is innocuous because z_t can include non-linear transformations of a set of more basic state variables y_t. The linear portfolio weights can be interpreted as more general portfolio weight functions $x_t = g(y_t)$ for any $g(\cdot)$ that can be spanned by a polynomial expansion in the more basic state variables y_t.

The obvious appeal of our approach is its simplicity and the fact that all of the statistical techniques designed for the static mean–variance problem can be applied directly to the single- and multiperiod market timing problems. Naturally, this simplicity comes with drawbacks that are discussed and evaluated carefully in Brandt and Santa-Clara (2006). We also demonstrate in the chapter how our parametric portfolio weights relate to the more traditional approach of modeling returns and state variables with a VAR in logs (equation (2.16)). Finally, we provide an extensive empirical application.

4.1.2. Large-Scale Portfolio Choice with Parametric Weights

Our approach in Brandt et al. (2009) is similar, in that we parameterize the optimal portfolio weights, but is geared toward large-scale cross-sectional applications. Suppose that at each date t there are large number of N_t stocks in the investable universe. Each stock i has an excess return of $r_{i,t+1}$ from date t to $t+1$ and a vector of characteristics $y_{i,t}$ observed at date t. For example, the characteristics could be the market beta of the stock, the market capitalization of the stock, the book-to-market ratio of the stock, and the lagged 12-month return on the stock. The investor's problem is to choose the portfolio weights $x_{i,t}$ to maximize the expected utility of the portfolio return $r_{p,t+1} = \sum_{i=1}^{N_t} x_{i,t} r_{i,t+1}$.

We parameterize the optimal portfolio weights as a function of the characteristics:

$$x_{i,t} = \overline{x}_{i,t} + \frac{1}{N_t}\theta' \hat{y}_{i,t} \tag{4.13}$$

where $\overline{x}_{i,t}$ is the weight of stock i in a benchmark portfolio, θ is a vector of coefficients to be estimated, and $\hat{y}_{i,t}$ are the characteristics of stock i standardized cross-sectionally to have a zero mean and unit standard deviation across all stocks at date t. This particular parameterization captures the idea of active portfolio management relative to a performance benchmark. The intercept is the weight in the benchmark portfolio and the term $\theta' \hat{y}_{i,t}$ represents the deviations of the optimal portfolio from the benchmark. The characteristics are standardized for two reasons. First, the cross-sectional distribution of $\hat{y}_{i,t}$ is stationary through time, while that of $y_{i,t}$ can be nonstationary (depending on the characteristic). Second, the standardization implies that the cross-sectional average of $\theta' \hat{y}_{i,t}$ is zero, which means that the deviations of the optimal portfolio weights from the benchmark weights sum to zero, and that the optimal portfolio weights always sum to one. Finally, the term $1/N_t$ is a normalization that allows the portfolio weight function to be applied to an arbitrary number of stocks. Without this normalization, doubling the number of stocks without otherwise changing the cross-sectional distribution of the characteristics results in twice as aggressive allocations, although the investment opportunities are fundamentally unchanged.

The most important aspect of our parameterization is that the coefficients θ do not vary across assets or through time. Constant coefficients across assets implies that the portfolio policy only cares about the characteristics of the stocks, not the stocks themselves. The underlying economic idea is that the characteristics fully describe the stock for investment purposes. Constant coefficients through time means that the coefficients that maximize the investor's conditional expected utility at a given date are the same for all dates and therefore also maximize the investor's unconditional expected utility. This allows us to estimate θ by maximizing the sample analogue of the unconditional expected utility:

$$\begin{aligned} \max_{\theta} \frac{1}{T}\sum_{t=0}^{T-1} u(r_{p,t+1}) &= \frac{1}{T}\sum_{t=0}^{T-1} u\left(\sum_{i=1}^{N_t} x_{i,t} r_{i,t+1}\right) \\ &= \frac{1}{T}\sum_{t=0}^{T-1} u\left(\sum_{i=1}^{N_t}\left(\overline{x}_{i,t} + \frac{1}{N_t}\theta'\hat{y}_{i,t}\right) r_{i,t+1}\right), \end{aligned} \tag{4.14}$$

for some prespecified utility function (e.g., mean–variance, quadratic, or CRRA utility).

Our approach has several practical advantages. First, it allows us to optimize a portfolio with a very large number of stocks, as long as the dimensionality of the parameter vector is kept reasonably low. Second, but related, the optimal portfolio weights are less prone to error maximization and over-fitting because we optimize the entire portfolio by choosing only a few parameters. The optimized portfolio weights tend to be far less extreme than the portfolio weights resulting from a more standard plug-in approach. Third, our approach implicitly takes into account the dependence of expected returns, variances, covariances, and higher-order moments on the stock characteristics, to the extent that cross-sectional differences in these moments affect the expected utility of the portfolio returns.

We develop several extensions of our parametric portfolio weights approach in Brandt et al. (2009), including parameterizations that restrict the optimal portfolio weights to be nonnegative and nonlinear parameterizations that allow for interactions between characteristics (e.g., small stocks with high momentum). We also show how the idea of cross-sectionally parameterizing the optimal portfolio weights can be combined naturally with the idea of parametric market timing described earlier. In particular, to allow the impact of the characteristics on the optimal portfolio weights to vary through time as a function of the macroeconomic predictors z_t, we suggest the parameterization:

$$x_{i,t} = \overline{x}_{i,t} + \frac{1}{N_t}\,\theta'\left(z_t \otimes \hat{y}_{i,t}\right) \tag{4.15}$$

where $\otimes$ again denotes the Kronecker product of two vectors. As in the pure market timing case, the optimization problem can then be rewritten as a cross-sectionally parameterized portfolio choice for an augmented asset space with naively managed portfolios.

4.1.3. Nonparametric Portfolio Weights

Although parameterized portfolio weights overcome the dependence on return models, they still suffer from potential mispecification of the portfolio weight function. In Brandt (1999), I develop a nonparametric approach for estimating the optimal portfolio weights without explicitly modeling returns or portfolio weights, which can be used as a mispecification check. The idea of my nonparametric approach is to estimate the optimal portfolio weights from sample analogues of the FOCs or Euler equations (2.12). These Euler equations involve conditional expectations that cannot be conditioned down to unconditional expectations, because the portfolio weights solving the Euler equations are generally different across economic states and dates. Instead, I replace the conditional expectations with nonparametric regressions and then solve for the portfolio weights that satisfy the resulting sample analogs of the conditional Euler equations.

Consider a single-period portfolio choice. The optimal portfolio weights x_t are characterized by the conditional Euler equations $\mathrm{E}_t\left[u'\left(x_t{}' r_{t+1} + R_t^f\right) r_{t+1}\right] = 0$. Suppose the returns are i.i.d. so that the optimal portfolio weights are the same across all states. In that case, we can take unconditional expectations of the conditional Euler equations to obtain a set of unconditional Euler equations that characterize the optimal unconditional portfolio weights $x_t \equiv x$. Replacing these unconditional expectations with sample averages in the spirit of method of moments estimation yields the estimator:

$$\hat{x} = \left\{ x : \frac{1}{T} \sum_{t=1}^{T} u'\left(x' r_{t+1} + R_t^f\right) r_{t+1} = 0 \right\}. \tag{4.16}$$

The same logic applies to a time-varying return distribution, except that the Euler equations cannot be conditioned down because the optimal portfolio weights depend on the macroeconomic state variables z_t (and/or firm characteristics $y_{i,t}$). Instead, we can directly replace the conditional expectations with sample analogs, where the sample analog of a conditional expectation is a locally weighted (in state-space) sample average. For a given state realization $z_t = z$, the resulting estimator of the optimal portfolio weights is

$$\hat{x}(z) = \left\{ x : \frac{1}{T h_T^K} \sum_{t=1}^{T} \omega\left(\frac{z_t - z}{h_T}\right) u'\left(x' r_{t+1} + R_t^f\right) r_{t+1} = 0 \right\}, \tag{4.17}$$

Where $\omega(\cdot)$ is a kernel function that weights marginal utility realizations according to how similar the associated z_t is to the value z on which the expectations are conditioned, and h_T denotes a sequence of kernel bandwidths that tends to zero as T increases.[18] (The factor Th_T^K assures that the weighted average is not degenerate.) Applying Eq. (4.17) to all values of z, one value at a time, recovers state-by-state the optimal portfolio weights.

To better understand this estimator, we can interpret it in a more standard nonparametric regression framework. For any portfolio weights x, the weighted average represents a kernel regression of the marginal utility realizations on the state variables. With optimal bandwidths, this kernel regression is consistent, in that:

$$\frac{1}{T h_T^K} \sum_{t=1}^{T} \omega\left(\frac{z_t - z}{h_T}\right) u'\left(x' r_{t+1} + R_t^f\right) r_{t+1} \overset{T\to\infty}{\longrightarrow} \mathrm{E}\left[u'\left(x' r_{t+1} + R_t^f\right) r_{t+1} \middle| z_t = z \right]. \tag{4.18}$$

It follows that the portfolios weights that set to zero the nonparametric regressions converge to the portfolio weights that set to zero the corresponding conditional expectations.

[18]The kernel function must satisfy $\omega(u) = \prod_{i=1}^{K} k(u_i)$ with $\int k(u)\mathrm{d}u = 1$, $\int uk(u)\mathrm{d}u = 0$, and $\int u^2 k(u)\mathrm{d}u < \infty$. A common choice is a K-variate standard normal density with $k(u) = \exp\{-1/2u^2\}/\sqrt{2\pi}$. See Härdle (1990) or Altman (1992) for a more detailed discussion of kernel functions.

The estimator is developed in greater detail and for a more general multiperiod portfolio choice problem with intermediate consumption in Brandt (1999). I also discuss the optimal bandwidth choice, derive the asymptotics of the estimator (with optimal bandwidths, it is consistent and asymptotically Gaussian with a convergence rate of $\sqrt{Th_T^K}$), and examine its finite sample properties through Monte Carlo experiments. In Brandt (2003), I locally parameterize the portfolio weights to further improve the finite sample properties (in the spirit of the local polynomial regression approach of Fan, 1993).

Kernel regressions are not the only way to nonparametrically estimate optimal portfolio weights from conditional Euler equations. Another way is to flexibly parameterize the portfolio weights with polynomial expansions, condition down the Euler equations, and estimate the polynomial coefficients using a standard method of moments approach. Yet another way is to flexibly parameterize the conditional expectations and construct sample analogs of the conditional Euler equations through polynomial regressions. Irrespective of the method, however, all of these estimators are limited in practice by some form of the "curse of dimensionality." For kernel regressions, the curse of dimensionality refers to the fact that the rate of convergence of the estimator to its asymptotic distribution deteriorates exponentially with the number of regressors. For polynomial expansion methods, the number of terms in an expansion of fixed order increases exponentially. Realistically, the curse of dimensionality means that we cannot reliably implement nonparametric estimators with more than two predictors (given the usual quarterly or monthly postwar data).

In Aït-Sahalia and Brandt (2001), we propose an intuitive way to overcome the curse of dimensionality in a portfolio choice context. Borrowing from the idea of index regressions (Powell et al., 1989), we collapse the vector of state variables z_t into a single linear index $z_t'\beta$ and then implement the kernel regression approach described earlier with this univariate index. The index coefficients β are chosen such that the expected utility loss relative to the original problem is minimized. (Empirically, the expected utility loss turns out to be negligible in most cases). We interpret the relative magnitude and statistical significance of each index coefficient as a measure of how important the corresponding state variable is to the investor's portfolio choice. We then use this interpretation to single out the one or two most important predictors for a range of different preferences.

ACKNOWLEDGMENTS

I thank Pedro Santa-Clara, Luis Viceira, and an anonymous referee for their comments and suggestions.

REFERENCES

Aït-Sahalia, Y., and M.W. Brandt, 2001, Variable selection for portfolio choice, *Journal of Finance* 56, 1297–1351.

Aït-Sahalia, Y., and M.W. Brandt, 2007, Portfolio and consumption choice with option-implied state prices, working paper, Princeton University.

Akian, M., J.L. Menaldi, and A. Sulem, 1996, On an investment-consumption model with transaction costs, *SIAM Journal of Control and Optimization* 34, 329–364.

Alexander, G., and A. Baptista, 2002, Economic implications of using a mean-VaR model for portfolio selection: A comparison with mean-variance analysis, *Journal of Economic Dynamics and Control* 26, 1159–1193.

Altman, N.S., 1992, An introduction to kernel and nearest-neighbor nonparametric regression, *American Statistician* 46, 175–185.

Ameriks, J., and S.P. Zeldes, 2004, How do household portfolio shares vary with age? working paper, Columbia University.

Ang, A., and G. Bekaert, 2007, Stock return predictability: Is it there? *Review of Financial Studies* 20, 651–707.

Ang, A., G. Bekaert, and J. Liu, 2005, Why stocks may disappoint, *Journal of Financial Economics* 76(3), pages 471–508.

Avramov, D., 2002, Stock return predictability and model uncertainty, *Journal of Financial Economics* 64, 423–458.

Avramov, D., 2004, Stock return predictability and asset pricing models, *Review of Financial Studies* 17, 699–738.

Balcer, Y., and K. Judd, 1987, Effects of capital gains taxation on life-cylce investment and portfolio management, *Journal of Finance* 42, 743–761.

Balduzzi, P., and A. Lynch, 1999, Transaction costs and predictability: Some utility cost calculations, *Journal of Financial Economics* 52, 47–58.

Barberis, N., 2000, Investing for the long run when returns are predictable, *Journal of Finance* 55, 225–264.

Barro, R.J., and J.W. Friedman, 1977, On uncertain lifetimes, *Journal of Political Economy* 85, 843–849.

Basak, S., and A. Shapiro, 2001, Value-at-risk based risk management: Optimal policies and asset prices, *Review of Financial Studies* 14, 371–405.

Benartzi, S., and R.H. Thaler, 1995, Myopic loss aversion and the equity premium puzzle, *Quarterly Journal of Economics* 110, 73–92.

Best, M.J., and R. Grauer, 1991, On the sensitivity of mean-variance efficient portfolios to changes in asset means: Some analytical and computational results, *Review of Financial Studies* 4, 315–342.

Black, F., and M.S. Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* 81, 637–654.

Black, F., and R. Jones, 1987, Simplifying portfolio insurance, *Journal of Portfolio Management* 14, 48–51.

Black, F., and R. Litterman, 1992, Global portfolio optimization, *Financial Analysts Journal* 48, 28–43.

Bodie, Z., and D.B. Crane, 1997, Personal investing: Advice, theory, and evidence, *Financial Analyst Journal* 53, 13–23.

Box, G.E.P., and G.C. Tiao, 1973, *Bayesian Inference in Statistical Analysis* (John Wiley and Sons: New York, NY).

Brandt, M.W., 1999, Estimating portfolio and consumption choice: A conditional euler equations approach, *Journal of Finance* 54, 1609–1646.

Brandt, M.W., 2003, Hedging demands in hedging contingent claims, *Review of Economics and Statistics* 85, 119–140.

Brandt, M.W., A. Goyal, P. Santa-Clara, and J.R. Stroud, 2005, A simulation approach to dynamic portfolio choice with an application to learning about return predictability, *Review of Financial Studies* 18, 831–873.

Brandt, M.W., and P. Santa-Clara, 2006, Dynamic portfolio selection by augmenting the asset space, *Journal of Finance* 61, 2187–2217.

Brandt, M.W., P. Santa-Clara, and R. Valkanov, 2009, Parametric portfolio policies: Exploiting characteristics in the cross section of equity returns, *Review of Financial Studies* forthcoming.

Britten-Jones, M., 1999, The sampling error in estimates of mean-variance efficient portfolio weights, *Journal of Finance* 54, 655–671.

Brown, S., 1978, The portfolio choice problem: Comparison of certainty equivalent and optimal Bayes portfolios, *Communications in Statistics: Simulation and Computation* B7, 321–334.

Browne, S., 1999, Beating a moving target: Optimal portfolio strategies for outperforming a stochastic benchmark, *Finance and Stochastics* 3, 275–294.

Camerer, C.F., 1995, Individual decision making, in J.H. Kagel, and A.E. Roth, ed.: *Handbook of Experimental Economics*. pp. 587–703 (Princeton University Press: Princeton, NJ).

Campbell, J.Y., and J.F. Cocco, 2003, Household risk managament and optimal mortgage choice, *Quarterly Journal of Economics* 118, 1449–1494.

Campbell, J.Y., and L.M. Viceira, 1999, Consumption and portfolio decisions when expected returns are time varying, *Quarterly Journal of Economics* 114, 433–495.

Campbell, J.Y., and L.M. Viceira, 2002, *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors* (Oxford University Press: New York, NY).

Campbell, J.Y., and R.J. Shiller, 1988, The dividend price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* 1, 195–228.

Campbell, J.Y., Y.L. Chan, and L.M. Viceira, 2003, A multivariate model of strategic asset allocation, *Journal of Financial Economics* 67, 41–80.

Canner, N., N.G. Mankiw, and D.N. Weil, 1997, An asset allocation puzzle, *American Economic Review* 87, 181–191.

Chan, L., J. Karceski, and J. Lakonishok, 1999, On portfolio optimization: Forecasting covariances and choosing the risk model, *Review of Financial Studies* 12, 937–974.

Chan, Y.L., and L.M. Viceira, 2000, Asset allocation with endogenous labor income: The case of incomplete markets, working paper, Hong Kong University of Science and Technology.

Chen, N.F., R. Roll, and S.A. Ross, 1986, Economic forces and the stock market, *Journal of Business* 59, 383–403.

Cho, D., 2007, Uncertainty in second moments: Implications for portfolio allocation, working paper, SUNY at Buffalo.

Chopra, V.K., and W.T. Ziemba, 1993, The effects of errors in the means, variances, and covariances, *Journal of Portfolio Management* 19, 6–11.

Cocco, J.F., 2000, Hedging house price risk with incomplete markets, working paper, London Business School.

Cocco, J.F., 2005, Portfolio choice in the presence of housing, *Review of Financial Studies* 18, 535–567.

Cochrane, J.H., 1989, The sensitivity of tests of the intertemporal allocation of consumption to near-rational alternatives, *American Economic Review* 79, 319–337.

Connor, G., 1997, Sensible return forecasting for portfolio management, *Financial Analyst Journal* 53, 44–51.

Connor, G., and R. Korajczyk, 1988, Risk and return in an equilibrium APT: Application of a new test methodology, *Journal of Financial Economics* 21, 255–290.

Cox, J.C., and C.F. Huang, 1989, Optimum consumption and portfolio policies when asset prices follow a diffusion process, *Journal of Economic Theory* 49, 33–83.

Cox, J.C., and C.F. Huang, 1991, A variational problem occurring in financial economics, *Journal of Mathematical Economics* 20, 465–487.

Cox, J.C., S.A. Ross, and M. Rubinstein, 1979, Option pricing: A simplified approach, *Journal of Financial Economics* 7, 229–263.

Cremers, M., 2002, Stock return predictability: A Bayesian model selection perspective, *Review of Financial Studies* 15, 1223–1249.

Cuoco, D., H. He, and S. Issaenko, 2007, Optimal dynamic trading strategies with risk limits, *Operations Research* forthcoming.

Cvitanic, J., 2001, Theory of portfolio optimization in markets with frictions, in *Handbook of Mathematical Finance* (Cambridge University Press: Cambridge), Chapter 17.

Dammon, R., C. Spatt, and H. Zhang, 2001a, Diversification and capital gains taxes with multiple risky assets, working paper, Carnegie Mellon University.

Dammon, R., C. Spatt, and H. Zhang, 2001b, Optimal consumption and investment with capital gains taxes, *Review of Financial Studies* 14, 583–616.

Dammon, R., C. Spatt, and H. Zhang, 2004, Optimal asset location and allocation with taxable and tax-deferred investing, *Journal of Finance* 59, 999–1037.

Davis, M., and A. Norman, 1990, Portfolio selection with transaction costs, *Mathematics of Operations Research* 15, 676–713.

de Athayde, G.M., and R.G. Flores, 2004, A general solution to three-moments portfolio choice, *Journal of Economics Dynamics and Control* 28, 1335–1352.

DeMiguel, A.V., and R. Uppal, 2005, Portfolio investment with the exact tax basis via nonlinear programming, *Management Science* 51, 277–290.

Detemple, J., R. Garcia, and M. Rindisbacher, 2003, Monte Carlo method for optimal portfolios, *Journal of Finance* 58, 401–446.

Duffie, D., and T. Sun, 1990, Transaction costs and portfolio choice in a discrete-continuous time setting, *Journal of Economic Dynamics and Control* 14, 35–51.

Dybvig, P.H., 1995, Duesenberry's racheting of consumption: Optimal dynamic allocation and investment given intolerance for any decline in standard of living, *Review of Economic Studies* 62, 287–313.

Efron, B.F., and C.N. Morris, 1977, Stein's paradox in statistics, *Scientific American* 236, 119–127.

Elton, E.J., and M.J. Gruber, 1978, Taxes and portfolio composition, *Journal of Finance* 41, 751–762.

Epstein, L., and S. Zin, 1989, Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework, *Econometrica* 57, 937–969.
Fama, E., and K. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.
Fama, E.F., 1970, Multiperiod consumption-investment decisions, *American Economic Review* 60, 163–174.
Fama, E.F., and K.R. French, 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3–25.
Fan, J., 1993, Local linear regression smoothers and their minimax efficiencies, *Annals of Statistics* 21, 196–216.
Flavin, M., and T. Yamashita, 2002, Owner-occupied housing and the composition of the household portfolio over the life-cylce, *American Economic Review* 92, 345–362.
Frost, P.A., and J.E. Savarino, 1986, An empirical Bayes approach to efficient portfolio selection, *Journal of Financial and Quantitative Analysis* 21, 293–305.
Frost, P.A., and J.E. Savarino, 1988, For better performance: Constrain portfolio weights, *Journal of Portfolio Management* 15, 29–34.
Gallmeyer, M., R. Kaniel, and S. Tompaidis, 2006, Tax management strategies with multiple risky assets, *Journal of Financial Economcis* 80, 277–290.
Garlappi, L., L.V. Naik, and J. Slive, 2001, Portfolio selection with multiple assets and capital gains taxes, working paper, University of British Columbia.
George, E.I., and R.E. McCulloch, 1993, Variable selection via Gibbs sampling, *Journal of the American Statistical Association* 88, 881–889.
Gilboa, I., and D. Schmeidler, 1989, Maxmin expected utility with non-unique prior, *Journal of Mathematical Economics* 18, 141–153.
Gomes, F., 2005, Portfolio choice and trading volume with loss-averse investors, *Journal of Business* 78, 675–706.
Gomes, F., and A. Michaelides, 2003, Portfolio choice with internal habit formation: A life-cycle model with uninsurable labor income risk, *Review or Economic Dynamics* 6, 729–766.
Goyal, A., and I. Welch, 2003, Predicting the equity premium with dividend ratios, *Management Science* 49, 639–654.
Green, R., and B. Hollifield, 1992, When will mean-variance efficient portfolios be well-diversified? *Journal of Finance* 47, 1785–1809.
Grossman, S., and G. Laroque, 1991, Asset pricing and optimal portfolio choice in the presence of illiquid durable consumption goods, *Econometrica* 58, 25–51.
Grossman, S.J., and J. Vila, 1989, Portfolio insurance in complete markets: A note, *Journal of Business* 62, 473–476.
GSAM Quantitative Strategies Group, 2000, An introduction to the Black-Litterman asset allocation model, working paper, Goldman Sachs Asset Management.
Gul, F., 1991, A theory of dissapointment aversion, *Econometrica* 59, 667–686.
Härdle, W., 1990, *Applied Nonparametric Regression* (Cambridge University Press: New York, NY).
Hakansson, N.H., 1969, Optimal investment and consumption strategies under risk, an uncertain lifetime, and insurance, *International Economic Review* 10, 443–466.
Hansen, L.P., and T. Sargent, 1995, Discounted linear exponential quadratic gaussian control, *IEEE Transactions on Automatic Control* 40, 968–971.

Harrison, M.J., and D.M. Kreps, 1979, Martingales and arbitrage in multiperiod securities markets, *Journal of Economic Theory* 2, 381–408.

Harvey, C.R., J.C. Liechty, M.W. Liechty, and P. Müller, 2004, Portfolio selection with higher moments, working paper, Duke University.

He, H., and N.D. Pearson, 1991, Consumption and portfolio policies with incomplete markets and short-sale constraints: The infinite dimensional case, *Journal of Economic Theory* 54, 259–304.

Heaton, J., and D.J. Lucas, 1997, Market frictions, savings behavior, and portfolio choice, *Macroeconomic Dynamics* 1, 76–101.

Heaton, J., and D.J. Lucas, 2000, Portfolio choice and asset prices: The importance of entrepeneurial risk, *Journal of Finance* 55, 1163–1198.

Hu, X., 2005, Portfolio choice for homeowners, *Journal of Urban Economics* 58, 114–136.

Huang, J., 2008, Taxable and tax-deferred investing: A tax-arbitrage approach, *Review of Financial Studies* 21, 2173–2207.

Jagannathan, R., and N. Kocherlakota, 1996, Why should older people invest less in stocks than younger people? *FRB of Minneapolis Quarterly Review* 20, 11–23.

Jagannathan, R., and T. Ma, 2003, Risk reduction in large portfolios: Why imposing the wrong constraints helps, *Journal of Finance* 58, 1651–1683.

James, W., and C. Stein, 1961, Estimation with quadratic loss, *Proceedings of the Fourth Berkeley Symposium on Mathematics and Statistics* pp. 361–379.

Jobson, J.D., and B. Korkie, 1980, Estimation of Markowitz efficient portfolios, *Journal of the American Statistical Association* 75, 544–554.

Jobson, J.D., and B. Korkie, 1981, Putting Markowitz theory to work, *Journal of Portfolio Management* 7, 70–74.

Jobson, J.D., B. Korkie, and V. Ratti, 1979, Improved estimation for Markowitz portfolios using James-Stein type estimators, *Proceedings of the American Statistical Association, Business and Economic Statistics Section* pp. 279–284.

Jorion, P., 1986, Bayes-Stein estimation for portfolio analysis, *Journal of Financial and Quantitative Analysis* 21, 279–292.

Judd, K.L., 1998, *Numerical Methods in Economics* (MIT Press: Cambridge, MA).

Kahneman, D., and A. Tversky, 1979, Prospect theory: An analysis of decision under risk, *Econometrica* 47, 263–291.

Kan, R., and G. Zhou, 2007, Optimal portfolio choice with parameter uncertainty, *Journal of Financial and Quantitative Analysis* 42, 621–656.

Kandel, S., and R.F. Stambaugh, 1996, On the predictability of stock returns: An asset-allocation perspective, *Journal of Finance* 51, 385–424.

Kane, A., 1982, Skewness preferences and portfolio choice, *Journal of Financial and Quantitative Analysis* 17, 15–25.

Karatzas, I., and S.E. Shreve, 1991, *Brownian Motion and Stochastic Calculus* (Springer Verlag: New York, NY).

Kataoka, A., 1963, A stochastic programming model, *Econometrica* 31, 181–196.

Kim, T.S., and E. Omberg, 1996, Dynamic nonmyopic portfolio behavior, *Review of Financial Studies* 9, 141–161.

Klein, R.W., and V.S. Bawa, 1976, The effect of estimation risk on optimal portfolio choice, *Journal of Financial Economics* 3, 215–231.

Koo, H.K., 1998, Consumptyion and portfolio selection with labor income: A continuous time approach, *Mathematical Finance* 8, 49–65.

Kraus, A., and R.H. Litzenberger, 1976, Skewness preferences and the valuation of risky assets, *Journal of Finance* 31, 1085–1100.

Lax, Y., 2002, Habit formation and lifetime portfolio selection, working paper, Goldman Sachs Asset Management.

Ledoit, O., and M. Wolf, 2003, Improved estimation of the covariance matrix of returns with an application to portfolio selection, *Journal of Empirical Finance* 10, 603–621.

Ledoit, O., and M. Wolf, 2004, Honey, I shrunk the sample covariance matrix, *Journal of Portfolio Management* pp. 110–119.

Leland, H., 2001, Optimal portfolio implementation with transaction costs and capital gains taxes, working paper, Unversity of California, Berkeley.

Liu, H., 2004, Optimal consumption and investment with transaction costs and multiple risky assets, *Journal of Finance* 59, 289–338.

Liu, W.F., 2002, Savings and portfolio decision: Observable implications of Knightian uncertainty, working paper, University of Washington.

Longstaff, F., J. Liu, and J. Pan, 2003, Dynamic asset allocation with event risk, *Journal of Finance* 58, 231–259.

Lynch, A., and S. Tan, 2009, Multiple risky assets, transaction costs and return predictability: Allocation rules and implications for U.S. investors, *Journal of Financial and Quantitative Analysis* forthcoming.

Maenhout, P.J., 2004, Robust portfolio rules and asset pricing, *Review of Financial Studies* 17, 951–983.

Maenhout, P.J., 2006, Robust portfolio rules and detection-error probabilities for a mean-reverting risk premium, *Journal of Economic Theory* 128, 136–163.

Markowitz, H.M., 1952, Portfolio selection, *Journal of Finance* 7, 77–91.

Marx, D.L., and R.R. Hocking, 1977, Moments of certain functions of elements in the inverse wishart matrix, working paper, Texas A&M University.

Merton, R.C., 1969, Lifetime portfolio selection under uncertainty: The continuous time case, *Review of Economics and Statistics* 51, 247–257.

Merton, R.C., 1971, Optimum consumption and portfolio rules in a continuous-time model, *Journal of Economic Theory* 3, 373–413.

Merton, R.C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867–887.

Merton, R.C., 1975, Theory of finance from the perspective of continuous time, *Journal of Financial and Quantitative Analysis* 7, 1851–1872.

Michaud, R.O., 1989, The markowitz optimization enigma: Is optimized optimal?, *Financial Analyst Journal* 45, 31–42.

Nielsen, L.T., and M. Vassalou, 2006, The instantaneous capital market line, *Economic Theory* 28, 651–664.

Nigmatullin, E., 2003, Bayesian model averaging for moment conditions models, working paper, University of Wisconsin-Madison.

Pastor, L., 2000, Portfolio selection and asset pricing models, *Journal of Finance* 50, 179–223.

Pastor, L., and R.F. Stambaugh, 2000, Comparing asset pricing models: An investment perspective, *Journal of Financial Economics* 56, 335–381.

Pastor, L., and R.F. Stambaugh, 2002, Investing in equity mutual funds, *Journal of Financial Economics* 63, 351–380.

Perold, A., and W. Sharpe, 1988, Dynamic strategies for asset allocation, *Financial Analyst Journal* 44, 16–27.

Powell, J.L., J.H. Stock, and T.M. Stoker, 1989, Semiparametric estimation of index coefficients, *Econometrica* 57, 1403–1430.

Quiggin, J.C., 1982, A theory of anticipated utility, *Journal of Economic Behavior and Organization* 3, 323–343.

Roy, A., 1952, Safety first and the holding of assets, *Econometrica* 20, 431–439.

Samuelson, P.A., 1969, Lifetime portfolio selection by dynamic stochastic programming, *Review of Economics and Statistics* 51, 239–246.

Samuelson, P.A., 1989, A case at last for age-phased reduction in equity, *Proceedings of the National Academy of Sciences* 86, 9048–9051.

Schroder, M., and C. Skiadas, 1999, Optimal consumption and portfolio selection with stochastic differential utility, *Journal of Economic Theory* 89, 68–126.

Schroder, M., and C. Skiadas, 2002, An isomorphism between asset pricing models with and without linear habit formation, *Review of Financial Studies* 15, 1189–1221.

Scowcroft, A., and J. Sefton, 2003, Enhanced indexation, in S. Satchell, and A. Scowcroft, ed.: *Advances in portfolio construction and implementation*. pp. 95–124 (Butterworth-Henemann: New York, NY).

Sharpe, W., 1963, A simplified model for portfolio analysis, *Management Science* 9, 277–293.

Shefrin, H., and M. Statman, 2000, Behavioral portfolio theory, *Journal of Financial and Quantitative Analysis* 35, 127–151.

Shimko, D.C., 1999, *Finance in Continuous Time: A Primer* (Blackwell: Malden, MA).

Simaan, Y., 1993, Portfolio selection and asset pricing three parameter framework, *Management Science* 5, 568–577.

Steele, J.M., 2001, *Stochastic Calculus and Financial Applications* (Springer Verlag: New York, NY).

Telser, L.G., 1956, Saftey first and hedging, *Review of Economic Studies* 23, 1–16.

Tepla, L., 2001, Optimal investment with minimum performance constraints, *Journal of Economic Dynamics and Control* 25, 1629–1645.

Theil, H., and A. Goldberger, 1961, On pure and mixed estimation in economics, *International Economic Review* 2, 65–78.

Tu, J., and G. Zhou, 2004, Data-generating process uncertainty: What difference does it make in portfolio decisions? *Journal of Financial Economics* 72, 385–421.

Viceira, L.M., 2001, Optimal portfolio choice for long-horizon investors with nontradable labor income, *Journal of Finance* 56, 433–470.

Wachter, J.A., 2002, Portfolio and consumption decisions under mean-reverting returns: An exact solution for complete markets, *Journal of Financial and Quantitative Analysis* 37, 63–91.

Weil, P., 1989, The equity premium puzzle and the risk-free rate puzzle, *Journal of Monetary Economics* 24, 401–421.

Yao, R., and H.H. Zhang, 2005, Optimal consumption and portfolio choices with risky housing and borrowing constraints, *Review of Financial Studies* 18, 197–239.

Zellner, Z.A., and V.K. Chetty, 1965, Prediction and decision problems in regression models from the Bayesian point of view, *Journal of the American Statistical Association* 60, 608–615.