Foundations and Trends® in
Finance
Vol. 2, No. 2 (2006) 83–190

DOI: 10.1561/0500000015

Portfolio Performance Evaluation

George O. Aragon[1] **and Wayne E. Ferson**[2]

[1] *W.P. Carey School of Business, Arizona State University, PO Box 873906, Tempe, AZ (480) 965-5810, George.Aragon@asu.edu*

[2] *Marshall School of Business, University of Southern California, 701 Exposition Boulevard, Los Angeles, CA 90089-142, ferson@marshall.usc.edu, http://www2.bc.edu/∼fersonwa*

Abstract

This paper provides a review of the methods for measuring portfolio performance and the evidence on the performance of professionally managed investment portfolios. Traditional performance measures, strongly influenced by the Capital Asset Pricing Model of Sharpe (1964), were developed prior to 1990. We discuss some of the properties and important problems associated with these measures. We then review the more recent Conditional Performance Evaluation techniques, designed to allow for expected returns and risks that may vary over time, and thus addressing one major shortcoming of the traditional measures. We also discuss weight-based performance measures and the stochastic discount factor approach. We review the evidence that these newer measures have produced on selectivity and market timing ability for professional managed investment funds. The evidence includes equity style mutual funds, pension funds, asset allocation style funds, fixed income funds and hedge funds.

Keywords: Portfolio performance; mutual fund performance; hedge funds; managed portfolios.

1

Introduction

This is a good time for a review of the academic literature on evaluating portfolio performance, concentrating on professionally managed investment portfolios. While the literature goes back to before the 1960s, recent years have witnessed an explosion of new methods for performance evaluation and new evidence on the subject. We think that several forces have contributed to this renaissance. The demand for research on managed portfolio performance increased as mutual funds and related investment vehicles became more important to investors in the 1980s and 1990s. During this period, equity investment became widely popular, as 401(k) and other defined-contribution investment plans began to dominate defined-benefit plans in the United States. Under such plans, individuals make their own investment choices from a menu of employer-specified options. At the same time, "baby-boomers" reached an age where they had more money to invest, and new investment opportunities were developing for investors in Europe and Asia that increased the demand for professionally managed portfolio products. This period also witnessed an explosive growth in alternative investments, such as hedge funds and private equity vehicles.

While the demand for research on investment performance has increased, the cost of producing this research has declined. Early studies relied on proprietary or expensive commercial databases for their fund performance figures, or researchers collected data by hand from published paper volumes. In 1997, the Center for Research in Security Prices introduced the CRSP mutual fund database, compiled originally by Mark Carhart, into the academic research market. Starting in about 1994, several databases on hedge fund returns and characteristics became available to academic researchers. Of course, during the same period the costs of computing have declined dramatically. In response to an increased demand and lower costs of production, the supply of research on fund performance expanded dramatically.

This chapter provides a selective review of the methods for measuring portfolio performance and the evidence on the performance of professionally managed investment portfolios. As the relevant literature is vast and expanding quickly, a complete survey is virtually impossible. This one reflects its authors' interests, and no doubt, biases.

Chapter 2 reviews the classical measures of portfolio performance developed between about 1960 and 1990. Our review emphasizes a unifying theme. We measure the total performance by comparing the returns on the managed portfolio to the returns of an *Otherwise Equivalent* (our terminology) benchmark portfolio. This is a portfolio with the same risk and other relevant characteristics as the managed portfolio, but which does not reflect the manager's investment ability. A manager with investment ability generates higher returns than the otherwise equivalent alternative, at least before fees and costs are considered. We first present the traditional measures, then review the important problems and properties associated with these measures.

Early studies frequently attempt to distinguish security selection versus market timing abilities on the part of fund managers. Timing ability is the ability to use superior information about the future realizations of common factors that affect overall market returns. A manager with timing ability may alter the asset allocation between stocks and safe assets or among other broad asset classes. Selectivity refers to the use of security-specific information, such as the ability to pick winning stocks or bonds within an asset class. We develop this dichotomy and discuss the ability of various performance measures to capture it. This section closes with a review of the evidence for managed portfolio performance based on the traditional measures. This discussion touches on the issues of survivorship bias and persistence in performance, among other topics.

Chapter 3 discusses Conditional Performance Evaluation. Here, the idea is to measure performance accounting for the fact that the expected returns and risks for investing may vary over time depending on the state of the economy. An example motivates the approach. We then discuss simple modifications to the traditional measures that attempt to condition on the state of the economy by using lagged variables as instruments.

Chapter 4 discusses the Stochastic Discount Factor Approach to performance measurement. We show briefly how this is related to the traditional alpha, what advantages the approach may have, and some recent developments.

Chapter 5 summarizes and illustrates the main issues in implementing the performance measures using hypothetical numerical examples. The examples are from the perspective of a fund-of-fund that must evaluate the performance of a sample of hedge funds using historical returns data.

Chapter 6 presents a brief discussion of the measures for investment performance of fixed income funds and Chapter 7 discusses hedge funds. Research on these fund types is still in an early stage of development, and these types of funds seem to present unique challenges for measuring risk-adjusted performance and for interpreting performance measures.

In Chapter 8, we review the modern empirical evidence on fund performance, which begins in about 1995 when studies began to use the CRSP-Carhart mutual fund database. We review the evidence that conditional measures have produced, both on selectivity and market timing ability. Also, in the mid-1990s data on hedge fund returns and characteristics first became available to academic researchers. We include a review of the evidence on hedge-fund performance.

Chapter 9 provides tabular summaries of the historical evidence on the performance of mutual funds and hedge funds using actual data. We describe how this evidence is related to the classical question of the informational efficiency of the markets. The various performance measures are interpreted by using and referring back to the concepts developed earlier in the text and Chapter 10 is the conclusion.

2

Classical Measures of Portfolio Performance

This chapter provides an overview of the classical measures of risk-adjusted portfolio performance. We first describe the general logic that lies behind all of the measures, and then define the individual measures. We then discuss the theoretical properties of the measures in more detail. Finally, we look at empirical estimation of the measures on actual managed portfolios, and review the empirical evidence based on the classical measures.

2.1 The Measures: An Overview

The main idea in most of the classical measures of investment performance is quite simple. The measures essentially compare the return of a managed portfolio over some evaluation period to the return of a benchmark portfolio. The benchmark portfolio should represent a feasible investment alternative to the managed portfolio being evaluated. If the objective is to evaluate the investment ability of the portfolio manager or management company, as has typically been the case, the benchmark should represent an investment alternative that is equivalent, in all return-relevant aspects, to the managed portfolio being evaluated, except that it should not reflect the investment ability of the manager. Let us call such a portfolio an "*Otherwise Equivalent*" (OE) benchmark portfolio. An OE benchmark portfolio requires that all of the portfolio characteristics that imply differences in expected returns are the same for the fund being evaluated and for the benchmark. The problem in practice is to operationalize this idea. Most of the available measures of portfolio performance may be understood in terms of their definitions of the OE benchmark.

In order to operationalize the concept of an OE benchmark, it is necessary to have some model for what aspects of a portfolio should lead to higher or lower expected returns. That is, some *asset pricing model* is required.[1] For this reason, portfolio performance measures and asset pricing models are inextricably linked, and the development of portfolio performance measures in the literature mirrors the development of empirical asset pricing models. A brief overview of this parallel development should be useful.

Historically speaking, the earliest asset pricing models made relatively simple predictions about what it means for a benchmark to be OE to a managed portfolio. The Capital Asset Pricing Model of Sharpe (CAPM, 1964) implies that all investors should hold a broadly diversified "market portfolio," combined with safe assets or "cash," according to the investor's tastes for risk. It follows that an OE portfolio is a broadly diversified portfolio, combined with safe assets or cash, mixed to have the same market risk exposure, or "beta" coefficient as the fund. This is the logic of Jensen's (1968) alpha, which remains one of the most widely used measures of risk-adjusted performance. If alpha is positive the manager earns an abnormal return relative to the alternative of holding the benchmark portfolio strategy.

Early performance measures were sometimes crude in their treatment of investment costs and fees. Mutual funds charge expenses that get deducted from the net assets of the funds and sometimes additional "load fees," such as those paid to selling brokers, or transactions

[1] While a few studies have claimed the ability to bypass the need for an asset pricing model, we argue that a benchmark model is always implied.

fees that may be paid into the assets of the fund to compensate other shareholders for the costs of buying and selling the underlying assets. In addition, funds' trading costs represent a drain from the net assets of the fund. A manager may be able to generate higher returns than an OE benchmark before costs, yet after costs investors' returns may be below the benchmark. If a fund can beat the OE benchmark on an after cost basis, we say that the fund *adds value* for investors, to distinguish this situation from one where the manager has investment ability, but either extracts the rents from this ability in the form of fees and expenses, or dissipates it through trading costs. We will say that a manager has *investment ability* if the managed portfolio outperforms the OE portfolio on a before-cost basis.

Formal models of market timing ability were first developed in the 1980s, following the intuitive regression model of Treynor and Mazuy (1966). In the simplest example, a market timer has the ability to change the market exposure of the portfolio in anticipation of moves in the stock market. Before the market goes up, the timer takes on more market exposure and generates exaggerated returns. Before the market goes down, the timer moves into safe assets and minimizes losses. Merton and Henriksson (1981) model this behavior as like a put option on the market. A successful market timer can be seen as producing free, or "cheap" put options. The OE portfolio for evaluating a market timer is therefore a combination of the market portfolio, safe assets, and an option on the market portfolio.

Following the CAPM, empirical asset pricing in the 1970's began to explore models in which exposure to more than a single market risk factor determines expected returns. Merton (1973) and Long (1974) develop models where investors should not simply hold a broad market index and cash, but should also invest in "hedge portfolios" for other economically relevant risks, like interest rate changes and commodity price inflation. Some investors may care more about inflation or interest rate changes than others, so they should adjust their portfolios in different ways to address these concerns. For example, an older investor whose anticipated lifetime of labor income is relatively short, and who is concerned about future cost-of-living risks, may want a portfolio that will pay out better if inflation accelerates in the future.

A younger investor, less concerned about these issues, may not be satisfied with the lower expected return that inflation hedging implies. It follows from a model with additional hedge portfolios that an OE portfolio for performance evaluation should have the same exposures as the managed portfolio to be evaluated, not just with respect to the overall market, but also with respect to the other relevant risk factors.

A related asset pricing model is the Arbitrage Pricing model of Ross (APT, 1976), which allows for several risk factors to determine assets' expected returns. In the case of the APT the number of factors depends on the dimensionality of the pervasive, or irreducible common risks in the asset markets. According to this model an OE portfolio must have the same exposures to all of the pervasive risk factors.

The papers of Long and Merton suggested interest rates and inflation as risk factors, but their models did not fully specify what all the hedge portfolios should be, or how many there should be. The APT specifies the factors only in a loose statistical sense. This leaves it up to empirical research to identify the risk factors or hedge portfolios. Chen et al. (1986) empirically evaluated several likely economic factors, and Chen et al. (1987) used these in an evaluation of equity mutual funds. Connor and Korajczyk (1988) showed how to extract statistical factors from stock returns in a fashion theoretically consistent with the APT, and Connor and Korajczyk (1986) and Lehmann and Modest (1987) used statistical factors in models for mutual fund performance evaluation.

Current investment management practice typically assumes that the OE portfolio is defined by the fund manager's investment "style." Roughly, style refers to a subset of the investment universe in which a manager is constrained to operate, such as small capitalization stocks versus large stocks, or "value" versus "growth" firms. The style constraint may be a self-declared specialization, or it may be imposed on the manager by the firm. This leads to the idea of "style exposures," analogous to the risk exposures implied by the multiple-beta asset pricing models. In this approach the OE portfolio has the same style exposures as the portfolio to be evaluated. The style-based approach is reflected prominently in academic studies following Fama and French (1996) such as Carhart (1997), who evaluates mutual funds using four factors derived from empirically observed patterns in stock portfolio returns, observed to be related to the characteristics of the stocks.

Daniel et al. (1997) further refine style-based performance measures by examining the actual holdings of mutual funds, and measuring the characteristics of the stocks held by the fund. The characteristics include the market capitalization or size, a measure of value (the ratio of book value to market value), and the return over the previous year. For a given fund, the OE portfolio is formed by matching the characteristics of the portfolio held by the fund with "passive" portfolios constructed to have the same characteristics.

In some cases the style of a fund is captured using the returns of other managed portfolios in the same market sector. The OE portfolio is then a combination of a manager's peers' portfolios. With this approach the measured performance is a zero-sum game, as the average performance measured in the peer group must be zero. This approach can make it easy to control for costs and risks, to the extent that the portfolio and its peers are similar in these dimensions. In such cases, the performance differential can be a relatively clean measure of value added. However, finding truly comparable peer funds may be a challenge. Many empirical studies that use cross-sectional analysis implicitly adopt a peer group approach. For example, a regression slope coefficient of performance on past fund returns implicitly demeans, or subtracts the average returns from the right-hand side variable. Thus, the past performance is measured relative to the average for all funds used in the regression.

When the actual holdings of the fund to be evaluated are available, it is possible to apply so-called weight-based performance measures. These measures essentially examine the covariance of the manager's actual holdings, measured as proportions or portfolio weights, with the subsequent returns of the assets. The idea is that a manager who increases the fund's exposure to a security or asset class before it performs well, or who anticipates and avoids losers, has investment ability.

2.1.1 The Sharpe Ratio

Perhaps the simplest risk-adjusted performance measure is the Sharpe ratio, used by Sharpe (1966) to evaluate mutual fund performance.[2] The Sharpe ratio for a portfolio p is defined as:

$$\mathrm{SR}_p = E(r_p)/\sigma(r_p), \tag{2.1}$$

where $r_p \equiv R_p - R_f$ is the return of the portfolio p, net of the return, R_f, to a safe asset or cash and $\sigma(r_p)$ is the standard deviation or volatility of the portfolio excess return. The Sharpe ratio measures the degree to which a portfolio is able to yield a return in excess of the risk-free return to cash, per unit of risk. If we were to graph expected excess return against risk, measured by the "volatility," or standard deviation of the excess return, the Sharpe ratio would be the slope of a line from the origin through the point for portfolio p.[3] Any fixed portfolio that combines the fund with cash holdings would plot on the same line as the portfolio R_p itself. As a performance measure, the Sharpe ratio of the fund is compared with the Sharpe ratio of the OE benchmark. If the ratio is higher for the fund, it performs better than the benchmark.

The Sharpe ratio is traditionally thought to make the most sense when applied to an investor's total portfolio, as opposed to any particular fund that represents only a portion of the investor's portfolio. The assumption is that what the investor really cares about is the volatility of his or her total portfolio, and various components of the portfolio combine to determine that via diversification, depending on the correlations among the various components. If applied to a single fund in isolation, the Sharpe ratio ignores the correlation of the fund with the other investments in the portfolio, and so it may not correspond in any meaningful way to the desirability of the fund as an investment. If the Sharpe ratio of a fund is higher than that of the investor's total

[2] Sharpe (1966) referred to the measure as the "Reward-to-Variability" ratio.

[3] Sharpe (1966) used the term "volatility" to refer to the regression coefficient of the fund on a market index, and "variability" to refer to the standard deviation of return. In more modern terminology the regression coefficient is called the "beta" of the fund and volatility refers to the standard deviation of return. In this review, we will stick with the more modern use of these terms.

investment portfolio, we may still be able to conclude that the investor should be interested in the fund. However, if it is lower, we cannot draw any conclusions without knowing about the correlations.

The Sharpe ratio may also be inappropriate when returns are highly nonnormal. For example, Leland (1999) shows that it is important to consider higher moments of the distributions if the performance measure is to accurately capture an investor's utility function. Furthermore, if the returns distributions are highly skewed, such as when options may be traded, the Sharpe ratio can be misleading. Goetzmann et al. (2005) show that by selling put options at fair market prices one can generate very high Sharpe ratios without investment skill. They also give an example where a manager with forecasting skill can have a low Sharpe ratio.

Despite these limitations, the Sharpe ratio is used in practice as a measure of portfolio performance. The ratio remains important in empirical asset pricing as well, for it has a number of interesting properties whose descriptions are beyond the scope of this review. Sharpe (1992) provides an overview and retrospective.

2.1.2 Jensen's Alpha

Alpha is perhaps the most well-known of the classical measures of investment performance. Indeed, like "Coke™" the term alpha has become generic. Using the market portfolio of the CAPM to form the OE benchmark, Jensen (1968) advocated the original version, or Jensen's alpha. The most convenient way to define Jensen's alphas is as the intercept, α_p^J, in the following time series regression:

$$r_{p,t+1} = \alpha_p^J + \beta_p r_{m,t+1} + u_{pt+1}, \tag{2.2}$$

where $r_{p,t+1}$ is the return of the fund in excess of a short term "cash" instrument. Using the fact that the expected value of the regression error is zero, we see how Jensen's alpha represents the difference between the expected return of the fund and that of its OE portfolio strategy:

$$\alpha_p^J = E\{r_p\} - \beta_p E\{r_m\} = E\{R_p\} - [\beta_p E\{R_m\} + (1 - \beta_p)R_f]. \tag{2.3}$$

The second line expresses the alpha explicitly in terms of the expected returns and portfolio weights that define the OE portfolio strategy. The OE portfolio has weight β_p in the market index with return R_m, and weight $(1 - \beta_p)$ in the risk-free asset or cash, with return R_f.

Versions of Jensen's alpha remain the most popular performance measures in academic studies. However, the measure has some disadvantages. For example, alpha does not control for nonsystematic sources of risk that could matter to investors (e.g., Fama, 1972). We discuss additional problems and properties of alpha below.

2.1.3 The Treynor Ratio

One problem with Jensen's alpha arises when $\beta_p > 1$, as might occur for aggressive funds such as small capitalization stock funds, aggressive growth funds or technology funds. In these cases the OE portfolio strategy involves a negative weight on the safe asset. This implies borrowing at the same rate as the risk-free return. In empirical practice, a short-term Treasury bill rate is often used to represent the safe asset's return. Few investors can borrow as cheaply as the US Treasury, so the implied OE benchmark strategy is not typically feasible.

Treynor (1965) proposed a measure that penalizes the portfolio in proportion to the amount of leverage employed:

$$T_p = E\{r_p\}/\beta_p. \tag{2.4}$$

Like the Sharpe ratio, a higher value of the Treynor ratio suggests better performance. Unlike the Sharpe ratio, the excess return is normalized relative to the systematic risk or beta, not the total risk or volatility. From Equation (2.2), the Treynor ratio and Jensen's alpha are clearly related, as $T_p = \alpha_p^J/\beta_p + E\{r_m\}$. Thus, if $\alpha_p^J = 0$, $T_p = E\{r_m\}$. For a portfolio with a nonzero alpha, $T_p - E\{r_m\} = \alpha_p^J/\beta_p$ is invariant to the amount of borrowing or lending at the rate R_f. Thus, the measure does not rely on an OE portfolio strategy that may not be feasible.

2.1.4 The Treynor–Black Appraisal Ratio

Black and Treynor (1973) studied a situation where security selection ability implies expectations of nonzero Jensen's alphas for individual securities, or equivalently, the benchmark market return index in (2.3) is not mean–variance efficient given those expectations. They derive the mean–variance optimal portfolio in this case and show that the optimal deviations from the benchmark holdings for each security depend on the "Appraisal Ratio":

$$\mathrm{AR}_i = [\alpha_i/\sigma(u_i)]^2, \tag{2.5}$$

where $\sigma(u_i)$ is the standard deviation of the residual for security i in Equation (2.2). Taking the deviations from benchmark for all the securities defines the "active portfolio," and Black and Treynor show how the active portfolio's appraisal ratio depends on the ratios of the individual stocks. They note that the portfolio's appraisal ratio, unlike Jensen's alpha, is invariant both to the amount of benchmark risk and the amount of leverage used in the portfolio. Market timing ability may lead to changes in the market risk and leverage used. Black and Treynor argue that this invariance recommends the appraisal ratio as a measure of a portfolio manager's skill in gathering and using information specific to individual securities.

2.1.5 The Merton–Henriksson Market Timing Measure

Classical models of market-timing use convexity in the relation between the fund's return and the "market" return to indicate timing ability. In these models the manager observes a private signal about the future performance of the market and adjusts the market exposure or beta of the portfolio at the beginning of the period. Successful timing implies higher betas when the market subsequently goes up, or lower betas when it goes down, leading to the convex relation. In the model of Merton and Henriksson (1981), the manager shifts the portfolio weights discretely, and the resulting convexity may be modeled with put or call options. The Merton–Henriksson market timing regression is

$$r_{pt+1} = a_p + b_p r_{mt+1} + \Lambda_p \mathrm{Max}(r_{mt+1}, 0) + u_{t+1}. \tag{2.6}$$

The coefficient Λ_p measures the market timing ability. If $\Lambda_p = 0$, the regression reduces to the market model regression (2.2) used to measure Jensen's alpha. Thus, under the null hypothesis that there is no timing ability, the intercept measures performance as in the CAPM. However, if there is timing ability and Λ_p is not zero, the interpretation of the intercept a_p is different. The OE portfolio for this model and the interpretation of the intercept is more complex than in the case of Jensen's alpha, and we defer a discussion of this to Section 2.2.1.

2.1.6 The Treynor–Mazuy Market Timing Measure

The Treynor–Mazuy (1966) market-timing model is a quadratic regression:

$$r_{pt+1} = a_p + b_p r_{mt+1} + \Lambda_p r_{mt+1}^2 + v_{t+1}. \tag{2.7}$$

Treynor and Mazuy (1966) argue that $\Lambda_p > 0$ indicates market-timing ability. When the market is up the fund will be up by a disproportionate amount and when the market is down, the fund will be down by a lesser amount. Admati et al. (1986) formalize the model, showing how it can be derived from a timer's optimal portfolio weight, assuming normal distributions and managers with exponential utility functions. They show that the timing coefficient Λ_p is proportional to the product of the manager's risk tolerance and the precision of the signal about the future market returns. Admati et al. (1986) show how to separate the effects of risk aversion and signal quality by estimating regression (2.7) together with a regression for the squared residuals of (2.7), on the market excess return. Coggin et al. (1993) implemented this approach on equity mutual fund data.

2.1.7 Multibeta Models

Multibeta models arise when investors optimally hold combinations of a mean–variance efficient portfolio and hedge portfolios for the other relevant risks. In this case the OE portfolios are combinations of a mean–variance efficient portfolio and the relevant hedge portfolios. The simplest performance measures implied by multibeta models are straightforward generalizations of Jensen's alpha, estimated as the intercept in a multiple regression:

$$r_p = \alpha_p^M + \sum_{j=1,\ldots,K} \beta_{pj} r_j + v_p, \tag{2.8}$$

where the r_j, $j = 1, \ldots, K$, are the excess returns on the K hedge portfolios ($j = 1$ can be a market index). By the same logic used above in the CAPM, we can see that the performance measure α_p^M is the difference between the average excess return of the fund and the OE portfolio:

$$\begin{aligned}\alpha_p^M &= E\{r_p\} - \sum_j \beta_{pj} E\{r_j\} \\ &= E\{R_p\} - \left[\sum_j \beta_{pj} E\{R_j\} + (1 - \sum_j \beta_{pj}) R_f\right].\end{aligned} \tag{2.9}$$

2.1.8 Weight-Based Performance Measures

In a returns-based measure the return earned by the fund is compared with the return on the OE benchmark over the evaluation period. The OE benchmark is designed to control for risk, style, investment constraints and other factors, and the manager who returns more than the benchmark has a positive "alpha." One strength of returns-based methodologies is their minimal information requirements. One needs only returns on the managed portfolio and the OE benchmark. However, this ignores potentially useful information that is sometimes available in practice: the composition of the managed portfolio. In a weight-based measure the manager's choices are directly analyzed for evidence of superior ability. A manager who increases the fund's exposure to a security or asset class before it performs well, or who anticipates and avoids losers, is seen to have investment ability.

Cornell (1979) was among the first to propose using portfolio weights to measure the performance of trading strategies. Copeland and Mayers (1982) modify Cornell's measure and use it to analyze Value Line rankings. Grinblatt and Titman (1993) propose a weight-based measure of mutual-fund performance. Applications to mutual funds include Grinblatt and Titman (1989a), Grinblatt et al. (1995), Zheng (1999), Daniel et al. (1997) and others. The intuition behind weight-based performance measures can be motivated, following Grinblatt and Titman (1993), with a single-period model where an investor maximizes the expected utility of terminal wealth.

$$\text{Max}_x E\{U(W_0[R_f + x'r])|S\}, \tag{2.10}$$

where R_f is the risk-free rate, r is the vector of risky asset returns in excess of the risk-free rate, W_0 is the initial wealth, x is the vector of portfolio weights on the risky assets, and S is the manager's private information signal, available at time 0. Private information, by definition, is correlated with the future excess returns, r. If returns are conditionally normal given S, and the investor has nonincreasing absolute risk aversion, the first and second-order conditions for the maximization imply (see Grinblatt and Titman, 1993) that

$$E\{x(S)'[r - E(r)]\} = \text{Cov}\{x(S)'r\} > 0, \qquad (2.11)$$

where $x(S)$ is the optimal weight vector and $r - E(r)$ are the unexpected, or abnormal returns, from the perspective of an observer without the signal. Equation (2.11) says that the sum of the covariances between the weights of a manager with private information, S, and the abnormal returns for the securities in a portfolio is positive. If the manager has no private information, S, then the covariance is zero.

Grinblatt and Titman and others operationalize Equation (2.11) by introducing a set of benchmark weights, x_B. The empirical measure is

$$E\{[x(S) - x_B]'[r - E(r)]\} = \text{Cov}\{[x(S) - x_B]'r\}. \qquad (2.12)$$

The benchmark weights are introduced to improve the statistical properties of the estimated measures. (Although, the relevant statistical issues have yet to be studied.) Clearly, if the benchmark contains no special information about future returns, then $\text{Cov}\{x_B'r\} = 0$ and it does not change the measure in theory to introduce the benchmark weights. With the benchmark weights, the OE portfolio implied by the weight-based measure is given by $x_B'r$.

2.2 Properties of the Classical Measures

2.2.1 Selectivity vs. Timing

2.2.1.1 Merton–Henriksson Model

The Merton–Henriksson market timing regression is repeated for convenience:

$$r_{pt+1} = a_p + b_p r_{mt+1} + \Lambda_p \text{ Max}(r_{mt+1}, 0) + u_{t+1}. \qquad (2.6)$$

The coefficient Λ_p measures the market timing ability. If $\Lambda_p = 0$, the regression reduces to the market model regression (2.2) used to measure Jensen's alpha. Thus, under the null hypothesis that there is no timing ability, the intercept measures performance as in the CAPM. However, if Λ_p is not zero and there is timing ability, the interpretation is different.

In the model of Merton and Henriksson, the OE portfolio holds a weight of b_p in the market index returning R_m, a weight of $\Lambda_p P_0$ in an option with beginning-of-period price P_0 and return $\{\text{Max}(R_m - R_f, 0)/P_0 - 1\}$ at the end of the period, and a weight of $(1 - b_p - \Lambda_p P_0)$ in the safe asset returning R_f. The option is a one-period European call written on the relative value of the market index, $V_m/V_0 = 1 + R_m$, with strike price equal to the end of period value of one dollar invested in the safe asset, $1 + R_f$.

Given a measure of the option price P_0 it is possible to construct "timing adjusted" measures of performance. In practice, the price of the option must be estimated from an option pricing model. Let r_0 be the return on the option, measured in excess of the safe asset. The difference between the excess return of the fund and that of the OE portfolio may be computed as $\alpha_p = E(r_p) - b_p E(r_m) - \Lambda_p P_0 E(r_0)$. The measure α_p captures "total" performance in the following sense. If an investor holds the OE portfolio just described, he obtains the same beta and convexity with respect to the market as the fund. The difference between the fund's expected return and that of the OE portfolio reflects the manager's ability to deliver the same beta and convexity at a lower cost than in the market, and thus with a higher return. The essence of successful market timing is the ability to produce the convexity at low cost.

Note that the measure α_p is not the same as the intercept in regression (2.6). The intercept does not measure the fund's return in excess of the OE portfolio. Taking the expected value of (2.6), the intercept is related to the alpha as: $\alpha_p = a_p + \Lambda_p P_0 R_f$. The intercept, a_p, has been interpreted in the literature as "timing-adjusted" selectivity. This interpretation may be understood as follows. Consider a manager with "perfect" market timing, defined as the ability to obtain the option-like payoff at zero cost.[4] Such a manager would deliver the same payoff as the OE portfolio, while "saving" the cost of the option, $\Lambda_p P_0$. Increasing the position in the safe asset by this amount leaves the beta unchanged and produces the additional return, $\Lambda_p P_0 R_f$. The additional return is the difference between the intercept, a_p, and the alpha. If the manager can generate a higher return in excess of the OE portfolio than $\Lambda_p P_0 R_f$, the extra return is presumed to be attributable to selectivity. Under this interpretation, when $\alpha_p > \Lambda_p P_0 R_f$, then $a_p > 0$ measures the selectivity-related excess return, assuming perfect market timing ability.

The interpretation of a_p as timing adjusted selectivity breaks down if a manager has some timing ability, but the fund is not able to obtain the option-like payoff at zero cost. In this case, a_p reflects the imperfect timing ability, but also the effects of security selection, and the two are hard to disentangle. For example, a manager with timing ability who picks bad stocks may be hard to distinguish from a manager with no ability who buys options at the market price (e.g., Jagannathan and Korajczyk, 1986; Glosten and Jagannathan, 1994). Indeed, without an estimate of the option price P_0, the intercept a_p has no clean interpretation in general. In general, it is not possible to separate the effects of timing and selectivity on return performance without making strong assumptions about at least one of the components. But it is possible to measure the combined effects of timing and selectivity on total performance. The measure α_p captures the excess return of the fund over the OE portfolio, and thus, the total abnormal performance. Of course the measures of timing and total performance discussed in this section ignore trading costs, which in practice will reduce the return of the fund but not the OE portfolio when it is constructed using the usual market indexes. See Section ~~2.2.3~~ for more discussion of trading costs.[5]

[4] The term "perfect," as applied to market timing ability, has been used to refer to the special case where the option-like payoff is obtained at zero cost and also the coefficient $\Lambda_p = 1$ (e.g., Merton (1981, p. 369), Goetzmann et al. (2000, p. 259) We use the less restrictive definition given here.

[5] Aragon (2005~~, Chapter 2~~) translates α_p into a measure of value added and provides further discussion on the value of market-timing fund managers.

2.2.1.2 Treynor–Mazuy Model

Like the intercept of Equation (2.6), the intercept in the Treynor–Mazuy model of Equation (2.7) has been naively interpreted as a "timing adjusted" selectivity measure. However, as in the Merton–Henriksson model, the intercept does not capture the return in excess of an OE portfolio, because r_m^2 in this case, is not a portfolio return.[6] However, the model can be modified to capture the difference between the return of the fund and that of an OE portfolio. Let r_h be the excess return of the maximum correlation portfolio to the random variable r_m^2. A modified version of the model is the system:

$$\begin{aligned} r_p &= [\alpha_p + a_h \Lambda_p/\Lambda_h] + b_p r_m + \Lambda_p r_m^2 + \varepsilon_p, \\ r_h &= a_h + \Lambda_h r_m^2 + \varepsilon_h. \end{aligned} \tag{2.13}$$

In this system, easily estimated using the Generalized Method of Moments (Hansen, 1982), the parameter α_p is the difference between the average return of the fund and that of an OE portfolio consisting of the market index, the safe asset and the maximum correlation portfolio with excess return r_h. The idea is that the convexity in the fund's return obtained by its timing activities could be replicated by passively holding a properly leveraged position in the maximum correlation portfolio. A manager has investment ability if the fund's expected return exceeds that of the passive strategy with the same market beta and convexity.

The OE portfolio that replicates the beta and convexity of the fund has a weight of Λ_p/Λ_h in R_h and b_p in R_m, with $1 - b_p - \Lambda_p/\Lambda_h$ in the safe asset, assuming $\Lambda_h \neq 0$. If $\Lambda_h = 0$ the weight in R_h is set to zero. The fact that the OE portfolio has the same beta and convexity coefficient as r_p can be seen by substituting the regression for r_h on r_m^2 into the combination of r_m and r_h that defines the OE excess return. The means of r_p and the OE portfolio excess returns differ by $\alpha_p =$

[6] Note that the intercept in (2.7) can be interpreted as the difference between the fund's average return and that of a trading strategy that holds the market index and the safe asset, with a time-varying weight or beta in the market index equal to $b_p + \Lambda_p r_{mt+1}$. However, this weight is not feasible at time t without foreknowledge of the future market return, so this strategy is not a feasible OE portfolio. We consider OE portfolios that are feasible dynamic strategies in the discussion of conditional performance evaluation below.

$a_p - a_h\Lambda_p/\Lambda_h$, where a_p is the unrestricted intercept of (2.7). Thus, α_p measures the total return performance, in the presence of timing ability, on the assumption that timing ability may be captured by a quadratic function.

2.2.2 Differential Information

The very existence of investment ability presumes there is differential information: The portfolio manager needs to know more about future security returns than the average investor. Mayers and Rice (1979) study the question of whether a manager who knows more than "the market" as a whole would deliver a positive Jensen's alpha. They make the assumption that the manager with superior information has a small enough amount of capital that will have no effect on market prices. Thus, the information reflected in the market as a whole does not include the superior information. This is a strong assumption, perhaps more tenable at the time of their paper than today, given that institutional trading represents a larger fraction of trading volumes in recent years. Mayers and Rice provide assumptions under which the alpha of a manager with superior information is positive, as assessed by the investor without the information. They show that the manager, who better knows which states of the world are more likely, will invest more money in securities that pay off in the more likely states, while investing a smaller amount in the less likely states, thereby generating a higher average return than would be expected given the risk, as perceived by the uninformed investor. However, Mayers and Rice were not able to show that a positive alpha would be found under general assumptions.

Verrecchia (1980) showed that a positive alpha could be expected in the model of Mayers and Rice, if the fund manager maximized a utility function with constant relative or absolute risk aversion, but he also presented an example where a manager with a quadratic utility function would not deliver a positive alpha. A manager with quadratic utility will optimally ~~chose~~ a mean–variance efficient strategy, conditional on his information. Dybvig and Ross (1985) assumed that the informed manager uses his information to form a mean–variance efficient portfolio (conditionally efficient), and showed that when his returns are viewed by an investor without the information, they would not generally appear to be efficient (unconditionally mean–variance inefficient). Hansen and Richard (1987) showed more generally, that conditionally mean–variance efficient portfolios need not be unconditionally mean–variance efficient.

If a manager with superior information knows that he will be evaluated based on the unconditional mean and variance of the portfolio return, he may be induced to form a portfolio that uses his information to maximize the unconditional mean, relative to the unconditional variance. Ferson and Siegel (2001) derive such unconditionally efficient strategies and discuss their properties. They suggest that it might be useful to use the optimal unconditional efficient strategies, based on readily available public information, as a benchmark for portfolio evaluation. Ferson and Siegel (2006) and Aragon and Ferson (2008) further develop this idea.

While there are many studies of differential information, the literature has not fully resolved the effects of differential information in the problem of measuring portfolio performance. At some normative level this need not be of great concern. Given an OE portfolio that reflects a sensible alternative strategy to the portfolio being evaluated and controls for all of the managed portfolio's return-relevant characteristics, the alpha relative to the OE portfolio captures the extra return that the managed portfolio delivers. At a deeper level, however, it is disconcerting to think that a manager with investment ability in the form of superior information, could be using his information to maximize essentially the same objective as the investor/client, and yet be seen as delivering an inferior return from the client's perspective. Clearly, more research in this area is called for.

2.2.3 Nonlinear Payoffs

When managers may trade assets whose payoffs are nonlinearly related to the benchmark return, this may generate convexity or concavity in the fund returns. For example, Jagannathan and Korajczyk (1986) argue that a fund holding call options can appear to be a good market timer, based on the convexity of the fund's returns.

Managers trading options may also affect the fund's Sharpe Ratio and other measures of performance. For example, a strategy of selling out-of-the-money put options has little risk given small stock price changes, but has exaggerated risk given large stock price movements. Goetzmann et al. (2005) solve for the maximal Sharpe Ratio obtainable from trading options on the market index, and find that such strategies can generate large Sharpe Ratios. They suggest using the theoretically maximal Sharpe Ratios as a benchmark. Strategies that involve taking on default risk, liquidity risk, or other forms of catastrophic risk may generate upwardly biased Sharpe ratios.

Nonlinearities may also be more important for some kinds of investors and strategies. For example, some mutual funds face explicit constraints on derivatives use, whereas private investment firms such as hedge funds are not typically subject to such constraints. The problem of nonlinear payoffs for performance evaluation is more important for portfolios that have broad investment mandates, as these afford more opportunity to trade dynamically across asset classes. Dynamic trading is akin to the use of options or other derivatives in its ability to generate nonlinear payoffs. For example, in many option pricing models a derivative can be replicated by dynamic trading in the underlying assets. There are two cases to consider, when addressing the problem of nonlinear payoffs in a managed portfolio. In the first case, it may be possible to replicate the nonlinearity in the funds' returns by trading other assets whose returns are measured at the same frequency as the funds' returns. In this case, the OE portfolio includes the replicating securities, and the funds' ability may be measured as described above in our discussion of market timing models.

In the second case, it may not be possible to replicate the funds' nonlinearities with other security returns measured at the same frequency. This can occur for two reasons. First, the market may be "incomplete" with respect to the funds' returns: the patterns generated by the fund are just not available elsewhere. In this situation, it is necessary to use an asset pricing model based on an assumption about investors' utility functions. For example, Leland (1999) argues that a model featuring power utility is useful in such a context, as the nonlinearity of the marginal utility function captures the prices of nonlinear portfolio payoffs. The other situation occurs when the funds' nonlinearities can be replicated, but in order to do so it is necessary to trade at a higher frequency than the funds returns are measured. This situation leads to the problem of interim trading bias.

2.2.4 Interim Trading Bias

A potential cause of nonlinearity is "interim trading," which refers to a situation where fund managers trade and/or cash flows in and out of the fund within the period over which returns on the fund are measured. With monthly fund return data, a common choice in the literature, interim trading definitely occurs, as flows and trades occur typically each day. The problems posed by interim trading were perhaps first studied by Fama (1972), and later by Goetzmann et al. (2000), Ferson and Khang (2002) and Ferson et al. (2006a,b).

Consider an example where returns are measured over two periods, but the manager of a balanced fund trades each period. The manager has neutral performance, but the portfolio weights for the second period can react to public information at the middle date. By assumption, merely reacting to public information at the middle date does not require superior ability. You have to trade "smarter" than the general public to generate superior performance.

Suppose that a terrorist event at the middle date increases market volatility in the second period, and the manager responds by moving part of the portfolio into cash at the middle date. The higher volatility may indicate that the expected return-to-risk tradeoff for stocks has become less favorable for the second period, so the optimal portfolio is now more conservative. If only two-period returns can be measured and evaluated, the manager's strategy would appear to have partially anticipated the higher volatility. The fund's two-period market exposure would reflect some average of the before- and after-event positions. Measured from the beginning of the first period, the portfolio would appear to partially "time" the volatility-increasing event because of the move into cash. A returns-based measure over the two periods will detect this as superior performance. Goetzmann et al. (2000)

address interim trading bias by simulating the multiperiod returns generated by the option to trade between return observation dates. Ferson et al. (2006a,b) use continuous-time models to address the problem, as described below.

A weight-based measure can avoid interim trading bias by examining the conditional covariance between the manager's weights at the beginning of the first period and the subsequent two-period returns. The ability of the manager to trade at the intervening period does not enter into the measure, and thus interim trading creates no bias. Of course, managers may engage in interim trading based on superior information to enhance performance, and a weight-based measure will not record these interim trading effects either. Thus, the cost of using a weight-based measure to avoid bias is a potential loss of power. Ferson and Khang (2002) evaluate these tradeoffs, and conclude that the conditional weight-based measure described below is attractive.

2.2.5 Accounting for Costs

It is important to recognize the role of costs in any comparison of a managed portfolio return with that of a performance benchmark. The costs of investing include direct transactions costs, fees paid for portfolio management services and to meet regulatory requirements, sales and marketing expenses, and income taxes. In most of the academic studies that use the classical measures, the OE benchmark strategy does not pay these costs. Mutual fund returns, in contrast, are measured net of all the expenses summarized in the funds' expense ratio and also the trading costs incurred by the fund, but not reported in the expense ratio. Some funds may charge additional sales charges (load fees), or redemption fees upon purchase or sale of the funds' shares. Funds purchased through a broker may involve additional brokerage charges. Measured fund returns do not account for these additional charges. The measured returns for other types of funds, such as pension funds and hedge funds, may reflect or exclude these and other costs. Measuring the managed portfolio's returns and the performance benchmark returns on a cost-equivalent basis can get complicated.

The most common situation in academic empirical studies using returns-based measures has the OE benchmark paying no costs, while the measured return of the mutual fund being evaluated is net of the funds' expenses and trading costs. In this case the performance measure is somewhere between a before-cost measure of investment ability and an after-cost measure of value added. The measure is perhaps best interpreted as a partial value added for a "no load" fund investor whose tax payments for returns on the fund and the OE benchmark would be the same, and when the costs of trading the OE portfolio are trivial. Roughly speaking, a manager whose performance just matches the benchmark has enough ability to cover his or her trading costs and fees. Any performance in excess of the benchmark is value added for this hypothetical investor. Clearly the literature would benefit from more careful treatment of costs.

With weight-based measures it is possible to construct hypothetical returns by applying the portfolio weights to measured returns on the securities held by the funds. The result is a measure of hypothetical before-cost fund returns, which may be more comparable with an OE benchmark that pays no costs. Thus, weight-based measures seem attractive for measuring investment ability, if not for capturing value added. Weight-based measures have various other advantages and disadvantages, as we discuss below.

It would seem that a logical next step for research on fund performance measures is to more carefully take account of the full range of costs and taxes associated with investing in funds. There are asset pricing models that make predictions about the return-relevant measures of transactions costs and taxes. The problem, however, is that the incidence of many of these costs is likely to be different for different investors. For example, a pension plan pays no income taxes on the dividends or capital gains generated by a portfolio, so the manager and the plan client may care little about the form in which the gains are earned. A college endowment fund typically pays no taxes either, but if constrained to spend out of the income component of return, the managers of the endowment may have a preference for the form of the return. An individual investor may be taxed more favorably on capital gains than on dividends, and the relative tax cost may depend on the investor's total income profile. This implies that for a given fund, the OE portfolio for one investor may not be the same as the OE portfolio for another investor, and different investors may view the performance of the same fund in different ways. This suggests that future performance measures that reflect costs may be constructed for a range of hypothetical investors with different OE portfolios.

2.3 The Evidence for Professionally Managed Portfolios Using Classical Measures

This section reviews the evidence on fund performance based on the classical measures, focusing mainly on the evidence prior to the advent of the CRSP mutual fund database in the mid-1990s. We discuss more modern evidence and the evidence from conditional performance evaluation in subsequent sections.

2.3.1 Data Issues

Historical returns of mutual funds are made available through the NSAR forms that are filed every semiannual period in compliance with the Investment Company Act of 1940. Form NSAR requires each registered investment company to report the month-end net asset values of fund shares for that semiannual period. Net asset values reflect the value of a fund share to investors after expenses resulting from management fee and trading costs, but before deduction of any load fees or payment of personal taxes. Net returns therefore reflect the change in net asset value plus any dividend or capital gain distributions over the performance period.

Early studies of managed portfolio performance faced a variety of data-related challenges. Many studies used mutual fund return data that were hand-collected from Wiesenberger's *Investment Companies*, Moody's Investor Services manuals, and other print media. The methods for computing reported returns were not standardized. Early studies of managed portfolio performance were therefore restricted to small samples over short sample periods. Often, data on funds that has ceased to operate were not available, raising the issue of survivor selection bias in the available returns data. When funds entered the databases they often brought a track record of historical returns with them. If these are included in a study, a potential back-fill bias is created.

Two important data advances occurred in the mid-1980s. First, historical mutual fund returns data became available in an electronic format provided by Morningstar, and later through the CRSP Survivorship–Bias Free Mutual Fund Database–currently the two most widely used mutual fund databases.[7] Second, the semi-annual portfolio holdings reported in Form NSAR became electronically available by a private data vendor CDA Investment Technologies, Inc. Since 1985, mutual funds have been required to report their stockholdings to the SEC in Form NSAR. In addition to the semi-annual reports, most funds voluntarily disclose their holdings more frequently, on a quarterly basis, to the CDA database. Recently, many funds voluntarily report holdings monthly to Morningstar. An explosion in the number of mutual funds studies using return-based measures and studies using weight-based measures occurred as researchers combined the machine-readable data with more efficient computing power.

2.3.2 Survivorship

Survivorship creates a number of potential problems affecting both the average levels of performance and the apparent persistence in performance. One obvious reason for a manager to leave a database is poor performance. To the extent that managers are dropped because of poor performance, the measured performance of the surviving managers is biased upwards. For example, Elton et al. (1996) find an average bias of 0.7%–0.9% per year in mutual fund return data (see also Brown and Goetzmann (1995), Malkiel (1995), Gruber (1996), Carhart (1997)).

Brown et al. (1992), Hendricks et al. (1997) and Brown and Goetzmann (1995) consider the effects of survivorship on performance per-

[7] Elton et al. (2001) compare the CRSP with Morningstar, and find differences in alpha in the order of −4.5 bp and −18.7 bp per year for large and small stocks, respectively, over the period 1979–1998. However, they find that this difference has become much smaller over time.

sistence under the simplifying assumptions that the expected returns of all managers are the same, but there are differences in variances, and that managers leave the database when their returns are relatively low. Under these assumptions survivorship is likely to induce a spurious "J-shaped" relation between future and past relative returns. In particular, past poor performers that survived are likely to have reversed their performance later. Brown and Goetzmann (1995) show that a J-shape can fail to occur with nontrivial probability, in a sample satisfying the assumptions of Hendricks et al. (1997) when there is correlation across the funds. Carhart et al. (2002) show that the effects of survival selection on persistence in performance depends on the birth and death process and can be quite complex.

2.3.3 Early Evidence

Sharpe (1966) studies the performance of 34 equity style mutual funds, using annual returns for 1954–1963, computing both the Sharpe and Treynor ratios to measure performance. The two performance measures have a rank correlation in his sample of 97.4%. Measured net of expenses the funds perform below the Dow Jones 30 stock index. The average Sharpe ratio of the funds is 0.633, while that of the Dow Jones is 0.677, and only 11 of the 34 funds have Sharpe ratios above that of the index. Adding back expenses, however, 19 of the 34 funds beat the Dow Jones. Sharpe concludes that while some funds may be choosing portfolios with a better risk-return profile than the index, investors are not realizing better returns after costs. He finds that expense ratios in 1953 have some predictive power for subsequent performance; their rank correlation with the future Sharpe ratios is −50%.

Jensen (1968) studies the performance of 115 open end mutual funds over the period 1945–1964. Data are obtained from Wiesenberger's *Investment Companies*. He finds an average net-of-expense alpha of −1.1% per year relative to the SP500 index. The distribution of the alphas across funds is skewed to the left, with 66% of funds having a negative alpha. Using data on expense ratios, Jensen (1968) calculates fund performance gross of expenses (but net of trading costs). He finds a higher, yet still negative average performance of −0.4% per year.

Jensen concludes that, in aggregate, the investment ability of mutual fund managers is not great enough to recover even brokerage commissions.

Carlson (1970) studies the performance of common stock, balanced, and income style mutual fund portfolios over the period 1948–1967. He finds lower Sharpe Ratios for all mutual fund portfolios than for the S&P 500 and the NYSE Composite indices. However, he also finds that the results vary depending on mutual fund style and choice of market index. Specifically, the net Sharpe ratios of common stock and balanced fund portfolios exceed that of the Dow Jones Industrial Average, although income mutual fund portfolios under-perform all market indices. Carlson (1970) also calculates the Jensen's alpha for individual funds in his sample. In contrast to Jensen (1968), Carlson (1970) finds that the average net alpha is a positive 0.6% per year, and that the distribution of alpha across funds is positively skewed. He concludes that care must be exercised in generalizing from the performance results of a specific mutual fund group and a specific market index.

Cornell (1979) develops a weight-based performance measure where the expected return on the OE benchmark is the portfolio weighted-average of expected returns on the underlying securities. A security's expected return is calculated as the mean realized return over an estimation window. Copeland and Mayers (1982) apply this methodology to evaluate the Value Line Investment Survey over the period 1965–1978. They construct five stock portfolios based upon the survey's ranks of individual stocks, and find negative abnormal returns for the bottom quintile portfolio over the 26-weeks following the publication of the survey. This is consistent with the idea that the stocks assigned the lowest ranking correspond to negative private information by Value Line publishers. However, their estimates are only significant when expected returns on the portfolio securities are calculated based on the market model.

2.3.3.1 Selectivity and Timing

Kon (1983) studies the timing performance of 37 mutual funds over the period 1960–1976. He estimates a two and three-stage regime switching model to calculate the dynamics of a fund's beta relative to the value-weighted CRSP market index. Kon's measure of timing performance is equal to the sample covariance between the fund's beta and the market return. This measure is derived by Fama (1972). Kon finds no evidence of significant timing performance. In fact, the majority (23) of the 37 sample funds display negative timing performance. He concludes that the allocation of effort and expense of market timing activities should be re-evaluated.

Henriksson (1984) studies the timing performance of 116 open-end mutual funds over the period 1968–1980. He employs the parametric test developed by Merton and Henriksson (1981) and discussed in Section 2.1.5. Consistent with Kon (1983), Henriksson finds no evidence of market timing ability. In fact, only three of the sample funds display significant positive timing ability, whereas 62% of the funds have negative timing coefficients. Henriksson also find evidence of a "disturbing" negative relation between estimates of timing ability and selection ability. For example, 49 of the 59 funds with positive alpha also had a negative timing coefficient. He concludes that the funds that earn superior returns from stock selection appear to have negative market timing ability.

Chang and Lewellen (1984) estimate the Henriksson and Merton (1981) market-timing model for a sample of 67 mutual funds over the period 1971–1979. They find little evidence of timing ability and, if anything, funds display greater portfolio betas in down markets than in up markets. Consistent with Henriksson's (1984) findings, they also find evidence of a negative relation between stock selection and timing skills. In fact, the only two funds with a significant positive intercept were among the three funds to display significant perverse negative timing ability. They conclude that the evidence supports the general conclusion reached in previous studies; namely, that mutual funds do not display market timing ability.

2.3.3.2 Persistence

Some studies find that the performance of mutual funds may be persistent. Persistence means that funds that perform relatively well or poorly in the past may be expected to do so again in the future. Obviously, persistent performance, if it exists, should be of practical interest to fund investors.

Sharpe (1966) conducts one of the first examinations of the question of persistence in fund performance. Sharpe finds that the standard deviations of funds returns have some persistence over time. The rank correlation between the 1944–1953 and the 1954–1963 periods is 53%. However, the Sharpe ratios have less persistence. Their rank correlation is only 36%, with a t-ratio of 1.88. Subsequent studies by Jensen (1969), Carlson (1970), Ippolito (1992), Grinblatt and Titman (1992), Goetzmann and Ibbotson (1994), Shukla and Trzcinka (1994), Hendricks et al. (1993, 1997), Brown and Goetzmann (1995), Malkiel (1995), Gruber (1996), Elton et al. (1996) and Carhart (1997) all find some evidence of persistence in mutual fund performance.

One form of persistence is short-term continuation, or "momentum" in the funds relative returns. The presence of such momentum would suggest that investors could obtain better returns by purchasing those funds that have recently performed well, and by avoiding those that have recently performed poorly. Some of the above-cited studies find evidence of continuation over several months after portfolios of high-relative-return funds are formed. However, much of this continuation seems to be explained by funds' holdings of momentum stocks (e.g., Carhart (1997), Grinblatt et al. (1995)).

Much of the empirical evidence on performance persistence for mutual funds suggests a positive relation between the future and past performance, but concentrated in the poorly performing funds. Poor performance may be persistent. This is not as expected under the hypothesis that persistence is a spurious result, at least in a simple model of survivorship bias. Brown and Goetzmann (1995), Carhart (1997) and Elton et al. (1996) find similar patterns of persistence in samples of mutual funds designed to avoid survivorship bias. To the extent that fund performance persists, it seems to be mainly the poor performers.

The evidence on persistence for pension fund managers is relatively sparse. Christopherson and Turner (1991) study pension managers and conclude that "alpha at one time is not predictable from alpha at a previous time." Lakonishok et al. (1992) find some persistence of the relative returns of pension funds for 2–3 year investment horizons. Christopherson et al. (1998a,b) study persistence with conditional models, as described below. Similar to the studies of mutual funds, they find some evidence of persistence, but it is concentrated in the poorly performing funds.

2.3.3.3 Fund Flows and Performance

Ippolito (1989) first observed that funds whose past returns were relatively high tended to attract relatively more new investment money over the next year. Further evidence that investor flows chase recent high returns was found by Sirri and Tufano (1998), Chevalier and Ellison (1997) and others. Del Guercio and Tkac (2002) found that mutual fund investors pay more attention to simple measures of relative return than to more complex measures like alpha, in directing their new money flows. If past performance does not actually predict future performance, then it would seem that mutual fund investors as a group are behaving strangely. Gruber (1996) forms portfolios of mutual funds, weighted according to their recent new money flows. He finds that the new money earns higher average returns than the old money invested in equity style funds. This "smart money" effect is subsequently confirmed by Zheng (1999).

2.3.3.4 Tournaments and Risk Shifting

Studies of the relation between flows and performance find an interesting nonlinear shape to the relation between past performance and subsequent new money flows. Funds with the highest returns on average realize the largest subsequent inflows of new money, while funds with performance below the median do not experience withdrawals of a similar magnitude. This nonlinearity creates an incentive for funds akin to that of a call option, even if the manager's compensation is a fixed fraction of the assets under management. Brown et al. (1996) argue that managers may respond to this incentive with a risk-shifting strategy, on the assumption that performance evaluation occurs at annual periods. They find that those funds that are performing relatively poor near the middle of the year seem to increase their risk in the last five months of the year, as if to maximize the value of the option-like payoff caused by fund flows. Funds whose performance is relatively high near the middle of the year seem to lower their risk, as if to "lock in" their position in the performance tournament.

Koski and Pontiff (1999) examine the use of derivatives by mutual funds, motivated in part by the idea that derivatives may be a cost-effective tool for risk management. They find evidence of risk-shifting similar to Brown et al. (1996), but little evidence that the risk shifting is related to the use of derivatives. Busse (2001) re-examines the evidence for risk shifting using daily data, and argues that the evidence for this behavior in the earlier studies is exaggerated by a bias, related to return autocorrelations, in estimating the standard errors of monthly returns. Using daily data he finds no evidence for risk shifting behavior. Goriaev et al. (2005) also find that the evidence for risk shifting is not robust.

3

Conditional Performance Evaluation

Traditional measures of risk-adjusted performance for mutual funds compare the average return of a fund with an OE benchmark designed to control for the fund's average risk. For example, Jensen's (1968) alpha is the difference between the return of a fund and a portfolio constructed from a market index and cash with fixed weights. The portfolio has the same average market exposure, or "beta" risk as the fund. The returns and beta risks are typically measured as averages over the evaluation period, and these averages are taken "unconditionally," or without regard to variations in the state of financial markets or the broader economy. One weakness of this unconditional approach relates to the likelihood of changes in the state of the economy. For example, if the evaluation period covers a bear market, but the period going forward is a bull market, the performance evaluation may not have much validity.

In the Conditional Performance Evaluation (CPE) approach, fund managers' risk exposures and the related market premiums are allowed to vary over time with the state of the economy. The state of the economy is measured using predetermined, public information variables. Provided that the estimation period covers both bull and bear markets, we can estimate expected risk and performance in each type of market. This way, knowing that we are now in a bull state of the market for example, we can estimate the fund's expected performance given a bull state.

Problems associated with variation over time in mutual fund risks and market returns have long been recognized (e.g., Jensen, 1972; Grant, 1977), but CPE draws an important distinction between variation that can be tracked with public information and variation due to private information on the state of the economy. CPE takes the view that a managed portfolio strategy that can be replicated using readily available public information should not be judged as having superior performance. For example, in a conditional approach, a mechanical market timing rule using lagged interest rate data is not a strategy that requires investment ability. Only managers who correctly use more information than is generally publicly available, are considered to have potentially superior investment ability. CPE is therefore consistent with a version of semi-strong-form market efficiency as described by Fama (1970). While market efficiency can motivate the null hypothesis that conditional alphas are zero, one need not ascribe to market efficiency to use CPE. By choosing the lagged variables, it is possible to set the hurdle for superior ability at any desired level of information.

In addition to the lagged state variables, CPE like any performance evaluation requires a choice of benchmark portfolios. The first measures used a broad equity index, motivated by the CAPM. Ferson and Schadt (1996) used a market index and also a multifactor benchmark for CPE. Current practice is more likely to use a benchmark representing the fund manager's investment style.

3.1 Motivation and Example

The appeal of CPE can be illustrated with the following highly stylized numerical example. Assume that there are two equally-likely states of the market as reflected in investors' expectations; say, a "Bull" state and a "Bear" state. In a Bull market, assume that the expected return of the S&P500 is 20%, and in a Bear[1] market, it is 10%. The

[1] This differs from the conventional definition of a bear market, which some consider to be a 20% decline off of a previous high.

risk-free return to cash is 5%. Assume that all investors share these views — the current state of expected market returns is common knowledge. In this case, assuming an efficient market, an investment strategy using as its only information the current state, will not yield abnormal returns.

Now, imagine a mutual fund which holds the S&P500 in a Bull market and holds cash in a Bear market. Consider the performance of this fund based on CPE and Sharpe's (1964) CAPM. Conditional on a Bull market, the beta of the fund is 1.0, the fund's expected return is 20%, equal to the S&P500, and the fund's conditional alpha is zero.[2] Conditional on a Bear market, the fund's beta is 0.0, the expected return of the fund is the risk-free return, 5%, and the conditional alpha is, again, zero. A conditional approach to performance evaluation correctly reports an alpha of zero in each state. This is essentially the null hypothesis of a CPE analysis.

By contrast, an unconditional approach to performance evaluation incorrectly reports a nonzero alpha for our hypothetical mutual fund. Without conditioning on the state, the returns of this fund would seem to be highly sensitive to the market return, and the unconditional beta of the fund[3] is 1.5. The unconditional expected return of the fund is

[2] The conditional alpha given a bull state, according to the CAPM, is the fund's conditional expected excess return over cash minus its conditional beta multiplied by the conditional expected market excess return over cash, which is equal to $(0.20 - 0.05) - 1(0.20 - 0.05) = 0$.

[3] The calculation is as follows. The unconditional beta is $\mathrm{Cov}(F, M)/\mathrm{Var}(M)$, where F is the fund return and M is the market return. The numerator is:

$$\begin{aligned}\mathrm{Cov}(F,M) &= E\{(F - E(F))(M - E(M))|\mathrm{Bull}\} \times \mathrm{Prob(Bull)}\\ &\quad + E\{(F - E(F))(M - E(M))|\mathrm{Bear}\} \times \mathrm{Prob(Bear)}\\ &= \{(0.20 - 0.125)(0.20 - 0.15)\} \times 0.5 + \{(0.05 - 0.125)(0.10 - 0.15)\} \times 0.5\\ &= 0.00375.\end{aligned}$$

The denominator is:

$$\begin{aligned}\mathrm{Var}(M) &= E\{(M - E(M))^2|\mathrm{Bull}\} \times \mathrm{Prob(Bull)}\\ &\quad + E\{(M - E(M))^2|\mathrm{Bear}\} \times \mathrm{Prob(Bear)}\\ &= \{(0.20 - 0.15)^2\} \times 0.5 + \{(0.10 - 0.15)^2\} \times 0.5\\ &= 0.0025.\end{aligned}$$

The beta is therefore $0.00375/0.0025 = 1.5$. Note that the unconditional beta is not the same as the average conditional beta, because the latter is 0.5 in this example.

0.5(0.20) + 0.5(0.05) = 0.125. The unconditional expected return of the S&P500 is 0.5(0.20) + 0.5(0.10) = 0.15, and the unconditional alpha of the fund is therefore: (0.125 − 0.05) − 1.5(0.15 − 0.05) = **−7.5%**. The unconditional approach leads to the mistaken conclusion that the manager has negative abnormal performance. But the manager's performance does not reflect poor investment choices or wasted resources, it merely reflects common variation over time in the fund's conditional risk exposure and the market premium. The traditional model over adjusts for market risk and assigns the manager a negative alpha. However, investors who have access to information about the economic state would not use the inflated risk exposure and would therefore not ascribe negative performance to the manager.

3.2 Conditional Alphas

Conditional alphas are developed as a natural generalization of the traditional, or unconditional alphas. In the CPE approach the risk adjustment for a bull market state may be different from that for a bear market state, if the fund's strategy implies different risk exposures in the different states. Let $r_{m,t+1}$ be the excess return on a market or benchmark index. For example, this could be the S&P 500, a "style" index such as "small cap growth," or a vector of excess returns if a multi-factor model is used.

The model proposed by Ferson and Schadt is:

$$r_{p,t+1} = \alpha_p + \beta_o r_{m,t+1} + \beta'[r_{m,t+1} \otimes Z_t] + u_{pt+1}, \qquad (3.1)$$

where $r_{p,t+1}$ is the return of the fund in excess of a short term "cash" instrument, and Z_t is the vector of lagged conditioning variables, in demeaned form. The symbol $\otimes$ denotes the kronecker product, or element-by-element multiplication when $r_{m,t+1}$ is a single market index. A special case of Equation (3.1) is the classical CAPM regression, where the terms involving Z_t are omitted. In this case, α_p is Jensen's (1968) alpha.

To see how the model in Equation (3.1) arises, consider a conditional "market model" regression allowing for a time-varying fund beta, $\beta(Z_t)$, that may depend on the public information, Z_t:

$$r_{p,t+1} = \alpha_p + \beta(Z_t)r_{m,t+1} + u_{pt+1}, \tag{3.2}$$

with $E(u_{p,t+1}|Z_t) = E\{u_{p,t+1}r_{mt+1}|Z_t) = 0$. Now assume that the time-varying beta can be modeled as a linear function: $\beta(Z_t) = \beta_o + \beta' Z_t$. The coefficient β_o is the average conditional beta of the fund (as Z is normalized to mean zero), and the term $\beta' Z_t$ captures the time-varying conditional beta. The assumption that the conditional beta is a linear function of the lagged instrument can be motivated by a Taylor series approximation, or by a model such as Admati et al. (1986), in which an optimizing agent would endogenously generate a linear conditional beta by trading assets with constant betas, using a linear portfolio weight function.

Substituting the expression for the conditional beta into Equation (3.2), the result is Equation (3.1). Note that since $E(Z_t) = 0$, it follows that:

$$E\{\beta'[r_{m,t+1} \otimes Z_t]\} = \mathrm{Cov}\{\beta(Z_t), r_{m,t+1}\} = \mathrm{Cov}\{\beta(Z_t), E(r_{m,t+1}|Z_t)\}, \tag{3.3}$$

where the second equality follows from representing $r_{m,t+1} = E(r_{m,t+1}|Z_t) + u_{m,t+1}$, with $\mathrm{Cov}\{u_{m,t+1}, \beta(Z_t)\} = 0$. Equation (3.3) says that the interaction terms $\beta'[r_{m,t+1} \otimes Z_t]$ in the regression (3.1) control for common movements in the fund's conditional beta and the conditional expected benchmark return. This was the cause of the "bias" in the unconditional alpha in the example we used to motivate CPE. The conditional alpha, α_p in (3.1) is thus measured net of the effects of these risk dynamics.

The OE portfolio in this CPE setting is the "naive" dynamic strategy, formed using the public information Z_t, that has the same time-varying conditional beta as the portfolio to be evaluated. This strategy has a weight at time t on the market index equal to $\beta_o + \beta' Z_t$, and $\{1 - \beta_o - \beta' Z_t\}$ is the weight in safe asset or cash. Using the same logic as before, Equation (3.1) implies that α_p in the Ferson and Schadt model is the difference between the unconditional expected return of the fund and that of the OE strategy.

Christopherson et al. (1998a,b) propose a refinement of (3.1) to allow for a time-varying conditional alpha:

$$r_{p,t+1} = \alpha_{p0} + \alpha_p' Z_t + \beta_o r_{m,t+1} + \beta'[r_{m,t+1} \otimes Z_t] + u_{pt+1}. \qquad (3.4)$$

In this model, $\alpha_{p0} + \alpha_p' Z_t$ measures the time-varying, conditional alpha, and the OE portfolio is the same as in the previous case, but the time-varying alpha is now the difference between the conditional expected return of the fund, given Z_t, and the conditional expected return of the OE portfolio strategy. This refinement of the model may have more power to detect abnormal performance if performance varies with the state of the economy. For example, if a manager generates a large alpha when the yield curve is steep, but a small or negative alpha when it is shallow, the average abnormal performance may be close enough to zero that it cannot be detected using regression (2.7). In such a case regression (3.4) could track the time-variation in alpha and record this as a nonzero coefficient α_p on the instrument for the term structure slope.[4]

3.3 Conditional Market Timing

The classical market-timing model of Treynor and Mazuy (1966) was presented in Equation (2.7). Treynor and Mazuy (1966) argued that a market-timing manager will generate a return that bears a convex relation to the market. However, a convex relation may arise for a number of other reasons. One of these is common time-variation in the fund's beta risk and the expected market risk premium, related to public information on the state of the economy. Ferson and Schadt propose a CPE version of the Treynor–Mazuy model to handle this situation:

$$r_{pt+1} = a_p + b_p r_{mt+1} + C_p'(Z_t r_{m,t+1}) + \Lambda_p r_{mt+1}^2 + w_{t+1}. \qquad (3.5)$$

[4] The regression (3.4) also has statistical advantages in the presence of lagged instruments that may be highly persistent regressors with high autocorrelation, as is often the case in practice. Ferson et al. (2007) show that by including the $\alpha_p' Z_t$ term, the regression delivers smaller spurious regression biases in the beta coefficients. They also warn, however, that the t-statistics for the time-varying alphas are likely to be biased in this case.

In Equation (3.5), the term $C_p'(Z_t r_{m,t+1})$ controls for common time-variation in the market risk premium and the fund's beta, just like it did in the regression (3.1). A manager who uses Z_t linearly to time the market has no conditional timing ability, and thus $\Lambda_p = 0$. The coefficient Λ_p measures the market timing ability based on information beyond that contained in Z_t.

Interpreting the intercept in (3.5) raises issues similar to those in the classical Treynor–Mazuy regression. As described above, a_p is not the difference between the fund's return and that of an OE portfolio. The model may be generalized, given a conditional maximum correlation portfolio to r_m^2, along the lines previously described. Such an analysis has not yet appeared in the literature.

Merton and Henriksson (1981) and Henriksson (1984) describe the alternative model of market timing in which the quadratic term is replaced by an option payoff, $\text{Max}(0, r_{m,t+1})$, as described earlier. Ferson and Schadt (1996) develop a conditional version of this model as well.

In theoretical market-timing models the timing coefficient is shown to depend on both the precision of the manager's market-timing signal and the manager's risk tolerance. For a given signal precision, a more risk tolerant manager will implement a more aggressive timing strategy, thus generating more convexity. Similarly, for a given risk tolerance a manager with a more precise timing signal will be more aggressive. Precision probably varies over time, as fund managers are likely to receive information of varying uncertainty about economic conditions at different times. Effective risk aversion may also vary over time, according to arguments describing mutual fund "tournaments" for new money flows (e.g, Brown et al., 1996), which may induce managers to take more risks when their performance is lagging and to be more conservative when they want to "lock in" favorable recent performance. Therefore, it seems likely that the timing coefficient which measures the convexity of a fund's conditional relation to the market is likely to vary over time. Ferson and Qian (2004) allow for such effects by letting the timing coefficient vary over time as a function of the state of the economy. Replacing the fixed timing coefficient above with $\Lambda_p = \Lambda_{0p} + \Lambda_{1p}' Z_t$ we arrive at a conditional timing model with time-varying performance:

$$r_{pt+1} = a_p + b_p r_{mt+1} + C_p'(Z_t r_{m,t+1}) + \Lambda_{0p} r_{mt+1}^2 + \Lambda_{1p}'(Z_t r_{m,t+1}^2) + w_{t+1}. \quad (3.6)$$

In this model the coefficient Λ_{1p} on the interaction term $(Z_t r_{m,t+1}^2)$ captures the variability in the managers timing ability, if any, over the states of the economy. By examining the significance of the coefficients in Λ_{1p}, it is easy to test the null hypothesis that the timing ability is fixed against the alternative hypothesis that timing ability varies with the economic state. Ferson and Qian (2004) find evidence that market timing ability varies with the economic state.

Becker et al. (1999) further develop conditional market-timing models. In addition to incorporating public information, their model features explicit performance benchmarks for measuring the relative performance of fund managers. In practice, performance benchmarks represent an important component of some fund managers' incentives, especially for hedge funds that incorporate explicit incentive fees. Even for mutual funds, Schultz (1996) reports that Vanguard included incentive-based provisions in 24 of 38 compensation contracts with external fund managers at that time. Elton et al. (2003) find that about 10% of the managers in a sample of US mutual funds are compensated according to incentive contracts. These contracts determine a manager's compensation by comparing fund performance to that of a benchmark portfolio. The incentive contracts induce a preference for portfolio return in excess of the benchmark. The model of Becker et al. refines the conditional market timing models of Ferson and Schadt (1996) in two ways. First, it allows for explicit, exogenous performance benchmarks. Second, it allows for the separate estimation of parameters for risk aversion and the quality of the market timing signal, conditional on the public information.

Starks (1987), Grinblatt and Titman (1989b) and Admati and Pfleiderer (1997) present models of incentive-based management contracts, focusing on agency problems between managers and investors. Chiu and Roley (1992) and Brown et al. (1996), among others, examine the behavior of fund managers when relative performance is important.

Heinkel and Stoughton (1995) present unconditional market-timing models with benchmark investors.

3.4 Conditional Weight-Based Measures

The weight-based performance measures discussed in Chapter 2 are unconditional, meaning they do not attempt to control for dynamic changes in expected returns and volatility. Like the classical returns-based performance measures, unconditional weight-based measures have problems handling return dynamics. It is known that unconditional weight-based measures can show performance when the manager targets stocks whose expected return and risk have risen temporarily (e.g., stocks subject to takeover or bankruptcy); when a manager exploits serial correlation in stock returns or return seasonalities; and when a manager gradually changes the risk of the portfolio over time, as in style drift.[5] These problems may be addressed using a conditional approach.

Ferson and Khang (2002) develop the *Conditional Weight-based Measure* of performance (CWM) and show that it has a number of advantages. Like other CPE approaches, the measure controls for changes in expected returns and volatility, as captured by a set of lagged economic variables or instruments. However, the CWM uses the information in both the lagged economic variables and the fund's portfolio weights.

The Conditional Weight Measure is the average of the conditional covariances between future returns and portfolio weight changes, summed across the securities held. It generalizes Equation (2.11) as follows:

$$\mathrm{CWM} = E\left\{\sum_j w_j(Z,S)[r_j - E(r_j|Z)]\right\}. \tag{3.7}$$

The symbol $w_j(Z,S)$ denotes the portfolio weight at the beginning of the period. The weights may depend on the public information, denoted by Z. The weights of a manager with superior information,

[5] See Grinblatt and Titman (1993) for a discussion.

denoted by S, may also depend on the superior information. Superior information, by definition, is any information that can be used to predict patterns in future returns that cannot be discerned from public information alone.

In Equation (3.7) the term $r_j - E(r_j|Z)$ denotes the unexpected, or abnormal future returns of the securities, indexed by j. Here, we define the abnormal return as the component of return not expected by an investor who only sees the public information Z at the beginning of the period. For example, if returns are measured over the first quarter, $E(r|Z)$ is the expected return for the first quarter based on public information about the economy as of the last trading day of the previous December. The sum of the covariances between the weights, measured at the end of December, and the subsequent abnormal returns for the securities in the first quarter, is positive for a manager with superior information, S. If the manager has no superior information, S, then the covariance is zero.

In practice, just as with the unconditional weight-based measures, it may be useful to introduce a benchmark with weights, w_{jb}, that are included in the public information set Z at the beginning of the quarter. The modified measure is

$$\mathrm{CWM} = E\left\{ \sum_j [w_j(Z,S) - w_{jb}][r_j - E(r_j|Z)] \right\}. \tag{3.8}$$

Ferson and Khang (2002) define the benchmark weights at the beginning of quarter t as the portfolios' actual weights lagged k periods, updating these with a buy-and-hold strategy. Thus, each manager's position, k quarters ago, defines his "personal" benchmark. The underlying model thus presumes that a manager with no investment ability follows a buy-and-hold strategy over the k quarters. A manager with investment ability changes the portfolio in order to beat a buy-and-hold strategy. The weight-based measures are in the units of an excess return of the fund over the benchmark. Because w_{jb} is assumed to be known given Z, it will not affect the CWM in theory. However, the benchmark weights will affect the statistical properties of the measure.

The benchmark weights are assumed to be public information at time t. However, the date when past weights become public information will depend on the circumstances. Mutual funds' portfolio weights become publicly available at least every six months, by law, although with a reporting lag. Many funds now report their holdings monthly, or even more frequently. In application to pension funds, one may view the public information as that available to pension plan sponsors. If a plan sponsor wishes to know the current holdings of a portfolio, a manager is likely to respond within days, if not hours. However, plan sponsors systematically examine holdings data on a less frequent basis. A lag of one quarter may be the most reasonable assumption; however, one could argue for a longer period. For example, a careful review of the holdings may take place on an annual basis with a more cursory review at quarterly reporting periods. The time when the weights can legitimately be called public information is therefore not clear. Thus, Ferson and Khang (2002) use various lags, k, to evaluate the sensitivity of the measures to this issue.

4

The Stochastic Discount Factor Approach

Modern asset pricing theory posits the existence of a *stochastic discount factor*, m_{t+1}, which is a scalar random variable, such that the following equation holds:

$$E(m_{t+1}R_{t+1} - \underline{1}|Z_t) = 0, \qquad (4.1)$$

where R_{t+1} is the vector of primitive asset gross returns (payoff divided by price), $\underline{1}$ is an N-vector of ones and Z_t denotes the public information set available at time t. Virtually all asset pricing models may be viewed as specifying a particular stochastic discount factor, m_{t+1}. The elements of the vector $m_{t+1}R_{t+1}$ may be viewed as "risk adjusted" gross returns. The returns are risk adjusted by "discounting" them, or multiplying by the discount factor, m_{t+1}, so that the expected "present value" per dollar invested is equal to one dollar. Thus, m_{t+1} is called a stochastic discount factor (SDF). We say that a SDF "prices" the assets if Equation (4.1) is satisfied.

Re-arranging Equation (4.1) reveals that the expected return is determined by the SDF model as:

$$E_t(R_{t+1}) = [E_t(m_{t+1})]^{-1} + \mathrm{Cov}_t\{-m_{t+1}/E_t(m_{t+1}); R_{t+1}\}, \qquad (4.2)$$

where $\text{Cov}_t\{\cdot,\cdot\}$ is the conditional covariance given the information at time t and $[E_t(m_{t+1})]^{-1}$ is the risk-free or expected "zero-beta" return, known at time t. Thus, predicted excess returns differ across funds in proportion to the conditional covariances of their returns with the SDF.

Chen and Knez (1996) were the first to develop SDF alphas for fund performance. For a given SDF we may define a fund's conditional SDF alpha following Chen and Knez (1996) and Farnsworth et al. (2002) as:

$$\alpha_{pt} \equiv E(m_{t+1}R_{p,t+1}|Z_t) - 1. \tag{4.3}$$

Consider an example where m_{t+1} is the intertemporal marginal rate of substitution for a representative investor: $m_{t+1} = u'(C_{t+1})/u'(C_t)$, where $u'(C)$ is the marginal utility of consumption. In this case, Equation (4.1) is the Euler equation which must be satisfied in equilibrium. If the consumer has access to a fund for which the conditional alpha is not zero he or she will wish to adjust the portfolio, purchasing more of the fund if alpha is positive and less if alpha is negative.

The SDF alpha is the risk adjusted excess return on the fund, relative to that of a benchmark portfolio that is assumed to be correctly priced by the SDF. If R_{Bt+1} is the benchmark, then Equation (4.3) implies $\alpha_{pt} \equiv E(m_{t+1}[R_{p,t+1} - R_{Bt+1}]|Z_t)$.

The SDF alpha depends on the model for the SDF, and theoretically the SDF is not unique unless markets are complete. Thus, different SDFs can produce different measured performance. This mirrors the classical approaches to performance evaluation, where performance is sensitive to the chosen benchmark. For example, analogous to results from Roll (1978), for any SDF that prices the primitive assets there exists another SDF that also prices the primitive assets and re-orders the ranking of the nonzero performance measures across funds.

4.1 Relation to the Beta Pricing Approach

The SDF and traditional beta pricing methods are simply related, as the existence of an SDF that is linear in a set of factors is equivalent to the existence of a beta pricing model based on the same factors. In this case the difference between the two alphas is simply a matter of scale. To illustrate, assume that $m_{t+1} = a + br_{mt+1}$. Substituting into Equation (4.3), and expanding the expectation of the product into the product of the expectations and rearranging, we find that the SDF alpha in (4.3) is equal to $E(m)$ times the alpha in the beta pricing model (see Ferson (1995, 2003) for more discussion).

The unconditional SDF alpha that is formed ignoring Z_t is the unconditional mean of the conditional alphas, where the expectation is taken across the states. In this respect SDF alphas differ from beta pricing model alphas. The conditional SDF alpha given Z is $\alpha(Z) = E(mR - 1|Z)$ and the unconditional alpha is $\alpha_u = E(mR - 1)$, so $E(\alpha(Z)) = \alpha_u$. The conditional alpha of a beta pricing model, in contrast, is the SDF alpha divided by the risk-free rate. When the risk-free rate is time varying, Jensen's inequality implies that the expected value of the conditional alpha in the beta pricing model is not the unconditional alpha. Ferson and Schadt (1996) find that average conditional alphas and unconditional alphas from beta pricing models can differ empirically for equity style funds.

4.2 Estimation of SDF Alphas

While the conditional SDF alpha, α_{pt}, is in general a function of Z_t, it is simpler to discuss the estimation of $\alpha_p = E(\alpha_{pt})$. A useful approach for estimating unconditional SDF alphas is to form a system of equations as follows:

$$\begin{aligned} u1_t &= [m_{t+1}R_{t+1} - \underline{1}] \otimes Z_t \\ u2_t &= \alpha_p - m_{t+1}R_{p,t+1} + 1. \end{aligned} \tag{4.4}$$

The sample moment condition is $g = T^{-1}\sum_t (u1'_t, u2'_t)$. We can use the Generalized Method of Moments (Hansen, 1982) to simultaneously estimate the parameters of the SDF model and the fund's SDF alpha.

The system (4.4) may also be estimated using a two-step approach, where the parameters of the model for m_{t+1} are estimated in the first step and the fitted SDF is used to estimate alphas in the second step. Farnsworth et al. (2002) find that simultaneous estimation is dramatically more efficient. However, a potential problem with the simultaneous approach is that the number of moment conditions grows substantially if many funds are to be evaluated, and there are more funds than months in most studies.

Fortunately, Farnsworth et al. (2002) show that we can estimate the joint system (4.4) separately for each fund without loss of generality. Estimating a version of system (4.4) for one fund at a time is equivalent to estimating a system with many funds simultaneously. The estimates of α_p and the standard errors for any subset of funds are invariant to the presence of another subset of funds in the system.

5

Implementing the Measures: A Fund-of-Funds Perspective

This chapter provides a hypothetical application of the techniques used to evaluate a given set of portfolios. The main goal is to illustrate the required steps for avoiding the pitfalls associated with data biases, performing the statistical tests, and interpreting the results. The application is developed from the perspective of a funds-of-funds (FOF) hedge fund manager. Evaluating the performance of hedge funds is a major responsibility for FOF managers. This task is complicated by the voluntary reporting of historical returns and the lack of transparency about the underlying strategies.

5.1 Evaluating a Set of Individual Hedge Funds

Imagine that in early 2006, an FOF manager is invited to invest with the following three hedge funds: Lucky Capital, Merger LP, and Murky LLC. These funds have passed an initial due diligence screen that includes background checks, manager interviews, and other qualitative analysis. Available data for these funds include the organizational characteristics and historical returns reported to a hedge fund database. The characteristics for three actual funds (aside from a relabeling of fund names) are reported in Table 5.1.

Table 5.1 Fund characteristics at the end of 2005.

Name	Lucky Capital	Merger LP	Murky LLC
Primary style	Long/Short Equity	Event Driven	Fixed Income Arbitrage
Inception date	September 21, 1993	January 31, 1986	November 30, 1993
Performance start date	October 31, 1993	January 31, 1986	November 30, 1993
Date added to database	September 30, 2001	September 30, 2001	July 19, 1995
Minimum investment ($$)	5,000,000	100,000	250,000
Management fee (percent)	1.5	1.5	1
Incentive fee (percent)	20	20	15
High-watermark	1	0	0
Subscription frequency	Monthly	Monthly	Monthly
Redemption frequency	Monthly	Monthly	Quarterly
Redemption notice (days)	60	10	45
Lockup period (months)	6	0	0
Domicile country	Virgin Islands (British)	Virgin Islands (British)	Jersey
Fund identifier	478	464	1030

Note: The table reports organizational characteristics for each hedge fund. Data are available from the database at the end of 2005.

For example, Lucky Capital is a Long/Short Equity hedge fund that was founded on September 21, 1993. Its performance data begin in the following month, resulting in 147 monthly return observations at the end of 2005. Other data about the fund are also provided. For example, Lucky Capital is an offshore fund with a six month lockup, a minimum investment amount of $5 million, a 1.5% management fee, and a 20% performance ("incentive") fee subject to a high-water mark. The high-water mark field is a one or a zero, indicating the presence or absence of a high water mark.

The monthly net returns have already been reduced by the management and performance fees paid by investors to the fund manager, and by the costs of trading. Table 5.2 presents the annual net returns for each fund, along with the value-weighted return on NYSE/AMEX stocks and the return on 30-day Treasury Bills. Although we show the

Table 5.2 Annual fund net returns over the sample period.

Year	Lucky	Merger	Murky	Stocks	T-Bills
1986	—	0.05	—	0.18	0.06
1987	—	−0.03	—	0.05	0.05
1988	—	0.59	—	0.17	0.06
1989	—	0.09	—	0.31	0.08
1990	—	−0.02	—	−0.03	0.08
1991	—	0.16	—	0.31	0.06
1992	—	0.05	—	0.08	0.03
1993	0.06	0.30	0.02	0.10	0.03
1994	0.06	0.04	−0.12	0.01	0.04
1995	0.53	0.34	0.05	0.38	0.06
1996	0.25	0.10	0.11	0.23	0.05
1997	0.08	0.20	0.12	0.34	0.05
1998	0.15	0.08	−0.05	0.29	0.05
1999	1.00	0.11	0.15	0.21	0.05
2000	0.34	0.12	0.07	−0.08	0.06
2001	0.04	−0.01	0.05	−0.12	0.04
2002	−0.24	−0.01	0.04	−0.22	0.02
2003	0.34	0.03	0.08	0.29	0.01
2004	−0.05	0.04	0.03	0.11	0.01
2005	0.08	0.08	0.05	0.05	0.03

Note: The table reports annualized net fund returns computed from the raw monthly return series. Annual returns are also reported for the value-weighted return on NYSE/AMEX stocks, and for the return on 30-day Treasury Bills. Shaded cells correspond to fund return observations that are not backfilled.

annual figures in the table, the actual analysis is based on the monthly returns. We want to use all of the available data.

A careful performance measurement for these funds must account for several important features of the hedge fund industry. First, the voluntary nature of the reported data suggests that a fund might choose to report only the most optimistic data available. For example, Table 5.1 shows that Lucky Capital was added to the database in September 2001. However, there may be other, less successful funds managed by the same advisor that will never be reported to the database, thereby upward-biasing the inferences about an advisor's skill. It would be wise, therefore, for the FOF to distinguish between the full sample of returns and the sub-sample of returns (the shaded cells) that follow the date the fund decided to report. Section 6.1.2 provides a detailed discussion of the "backfilling bias."

The FOF must also account for the fact that a hedge fund's systematic risk exposure might change within the evaluation period, either through derivatives usage or dynamic trading strategies. For example, Merger LP reports Event Driven as its primary category. This category includes the merger arbitrage strategy, in which the shares of a merger target are purchased with the proceeds from selling short the shares of the proposed acquirer. Mitchell and Pulvino (2001) show that the returns from merger arbitrage are similar to the returns from writing put options on the stock market, because merger deals are less likely to succeed in depreciating markets. Chapter 3 discusses performance measures that accommodate time-varying systematic risk exposures.

Finally, a fund's illiquid asset holdings make it difficult for the fund to produce reliable estimates of net asset value (NAV). For example, the primary category of Murky LLC involves fixed income securities. If some of the debt securities are rarely traded, the available price quotes will be stale, thereby introducing a downward bias in estimates of a fund's systematic risk. Asness et al. (2001) show that stale NAV can have an important effect on estimates of hedge fund performance.

As an illustration, the FOF can use the following fund-level regression model to account for the backfilling bias, dynamic risk, and illiquid asset exposure:

$$\begin{aligned} r_{pt+1} = {} & a_p + b_p r_{mt+1} + b_{1p} r_{mt} + b_{2p} r_{mt-1} + b_{3p} r_{mt-2} \\ & + \Lambda_p \mathrm{Max}(r_{mt+1}, 0) + u_{t+1}, \end{aligned} \tag{5.1}$$

where $E[u_{pt}] = E[u_{pt}(r_{mt}, r_{mt-1}, r_{mt-2}, r_{mt-3})] = 0$. The above model nests the traditional Jensen's alpha model in which $\Lambda_p = b_{1p} = b_{2p} = b_{3p} = 0$. The backfilling bias can be addressed by comparing the estimate of the Jensen's alpha using the full sample of returns (Model 1) with the alpha estimate after excluding the monthly returns that precede the date the fund was added to the database (Model 2). The above equation also nests ~~to~~ the Merton–Henriksson (1981) market-timing model (Model 3), in which $b_{1p} = b_{2p} = b_{3p} = 0$. The option payoff in the timing model allows the fund's beta to change based upon market performance, and therefore allows for a specific form of dynamic risk exposure. Finally, the lagged market model (Model 4) advocated by Scholes and Williams (1977) is nested within the constraint that $\Lambda_p = 0$. Lagged market returns will have explanatory power to the extent that a fund's reported returns do not fully reflect the true economic return in a given month. Models 3 and 4 are estimated without backfilled data. In this illustration we use the US stock market index to form the OE portfolio. This may be appropriate for Lucky Capital and Merger LP. However, it is less appropriate for Muddy LLC, which is a fixed income fund. We would run the analysis with some fixed income returns in place of the US stock index and see how the results are affected.

5.1.1 Lucky Capital

Table 5.3 shows that Lucky Capital has positive alpha for the full sample. Its monthly risk-adjusted performance, at 0.77%, is significant. However, Model 2 shows that the performance is negative and insignificant after excluding backfilled data. Therefore, the positive performance can be explained by a backfilling bias. The R-squared in Model 2 implies that the market model captures almost 70% of the variation in the fund's returns. Incorporating the nonlinear market excess return in Model 3 adds no explanatory power. If we found significance of the nonlinear term we would explore the fund's total performance in the

Table 5.3 The performance of lucky capital.

	Lucky Capital			
Coefficient	1	2	3	4
b_p	0.725	0.853	0.940	0.830
	(0.080)**	(0.100)**	(0.153)**	(0.076)**
Λ_p			−0.185	
			(0.304)	
b_{1p}				0.204
				(0.060)**
b_{2p}				−0.109
				(0.057)
b_{3p}				0.197
				(0.067)**
a_p	0.77	−0.12	0.16	−0.16
	(0.36)*	(0.32)	(0.49)	(0.29)
Observations	147	52	52	52
R-squared	0.34	0.69	0.69	0.77

Robust standard errors in parentheses.

*Significant at 5% level; **Significant at 1% level.

Note: The table reports the estimated coefficients from various market model regressions. The dependent variable is the monthly excess fund return. The intercept is reported in percent per month. Significance levels correspond to a two-tailed test of the null hypothesis that the coefficient equals zero.

presence of market timing ability, as described in Section 2.1. The significant one and three month lagged market returns in Model 4 suggests that there is some illiquidity in Lucky Capital's portfolio. However, the intercept, at −0.16% per month, is negative and insignificant. We conclude that Lucky Capital does not add value to investors.

5.1.2 Merger LP

Table 5.4 shows that Merger LP has insignificant systematic risk using the sample of non-backfilled returns (Model 2). The estimated performance is neutral — the intercept is 0.07% per month and insignificant. In addition, stale NAV does not appear to be an issue because the lagged observations of the market return have no explanatory power in the fund return regressions. In Model 3, we estimate a significant positive market beta ($b_p = 0.115$) in down markets, but a flat relation between Merger LP's return and the market return when the excess market return is positive. (The up-market beta is

Table 5.4 Performance of merger LP.

	Merger LP			
Coefficient	1	2	3	4
b_p	0.348	0.034	0.115	0.032
	(0.143)*	(0.030)	(0.056)*	(0.027)
Λ_p			−0.174	
			(0.081)*	
b_{1p}				0.021
				(0.023)
b_{2p}				0.041
				(0.023)
b_{3p}				0.015
				(0.016)
a_p	0.31	0.07	0.33	0.06
	(0.28)	(0.079)	(0.10)**	(0.08)
Observations	240	52	52	52
R-squared	0.18	0.06	0.21	0.20

Robust standard errors in parentheses.
*Significant at 5% level; **Significant at 1% level.
Note: The table reports the estimated coefficients from various market model regressions. The dependent variable is the monthly excess fund return. The intercept is reported in percent per month. Significance levels correspond to a two-tailed test of the null hypothesis that the coefficient equals zero.

$b_p + \Lambda_p = 0.115 - 0.174 = -0.059$.)[1] This is consistent with Mitchell and Pulvino's (2001) finding of a negative relation between deal failure and market returns. Accounting for this nonlinearity also reveals a positive intercept of 0.33% per month. The payoff of Merger LP resembles the payoff from writing put options on the market. The issue is whether the fund is able to achieve this payoff cheaper than an OE benchmark. Section 2.2.1 shows that the appropriate OE for Model 3 consists of cash, a weight of b_p in the market index, and a weight $\Lambda_p P_0$ in a one-period call option written on the gross return of the market index with a strike price equal to the gross return on the riskless rate, where P_0 denotes the option price.[2] The difference between the excess return of the fund and that of the OE portfolio may be computed as $\alpha_p = a_p + \Lambda_p P_0 R_f$. Under Black and Scholes (1973), we can estimate P_0 using the average risk-free rate and sample volatility of returns on the market index. Using the sample standard deviation of monthly market returns of 4.1% and average monthly risk-free rate of 0.15%, this gives

$$\alpha_p = 0.33\% - 0.174 \times 0.0164 \times 1.0015 = 0.0442\%.$$

In other words, Merge LP delivers an additional 4.4 basis points per month relative to a stock/options option strategy that replicates the dynamic risk exposure of Merger fund's portfolio.

5.1.3 Murky LLC

Table 5.5 shows that Murky LLC has very little market exposure (market beta of 0.087 or 0.067) according to Model 1 or 2. Moreover, the table reveals a positive alpha of 0.19% per month. However, the R-squared is only 7%, suggesting that the Jensen's alpha model is misspecified. Accounting for dynamic risk exposure in Model 3 also reveals a positive alpha, but the coefficient on the up-market variable

[1] Note that these are not conditional betas given public information, as discussed in Chapter 3, because the market return at time $t + 1$ is not known at time t. We do not illustrate the CPE approach here for simplicity.

[2] Equivalently, by put-call parity, the OE consists of cash, a weight of $b_p + \Lambda_p$~~P_0~~ in the market index, and put options on the gross market return with an exercise price equal to the gross return on the riskless asset.

Table 5.5 Performance of Murky LLC.

	Murky LLC			
Coefficient	1	2	3	4
b_p	0.087	0.067	0.165	0.069
	(0.043)*	(0.040)	(0.082)*	(0.039)
Λ_p			−0.201	
			(0.120)	
b_{1p}				0.055
				(0.027)*
b_{2p}				0.058
				(0.024)*
b_{3p}				0.012
				(0.020)
a_p	0.04	0.19	0.55	0.11
	(0.14)	(0.11)	(0.18)**	(0.11)
Observations	146	126	126	126
R-squared	0.06	0.07	0.13	0.17

Robust standard errors in parentheses.

*Significant at 5% level; **Significant at 1% level.

Note: The table reports the estimated coefficients from various market model regressions. The dependent variable is the monthly excess fund return. The intercept is reported in percent per month. Significance levels correspond to a two-tailed test of the null hypothesis that the coefficient equals zero.

is insignificant. Model 4 shows that lagged observations of the market index excess return are significant. Summing the betas reveals that the "true" market beta is approximately 0.18. An F-test shows that the change in R-squared from 0.07 to 0.17 is significant. More importantly, whereas in Model 2 we could reject the null hypothesis of zero intercept at the 7.5% significance level, Model 4 shows that the intercept, at 0.11% per month, is insignificant after accounting for a stale price bias in the fund's reported returns. We conclude that Murky LLC does not add value to investors.

6

Bond Fund Performance Measurement

6.1 Fixed Income Models

Elton et al. (1993; 1995) were the seminal academic studies of the performance of bond style mutual funds. They used versions of the classical multibeta model alphas described above, where the factors are selected to address the risks most likely to be important for fixed income portfolios.

Ferson et al. (2006a) brought modern term structure models to the problem of measuring bond fund performance. These models specify a continuous-time stochastic process for the underlying state variable(s). For example, let X be the state variable following a diffusion process:

$$\mathrm{d}X = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}w, \tag{6.1}$$

where $\mathrm{d}w$ is the local change in a standard Wiener process. The state variable may be the level of an interest rate, the slope of the term structure, etc. The model also specifies the form of a market price of risk, $q(X)$, associated with the state variable. The market price of risk is the expected return in excess of the instantaneous interest rate, per unit of state variable risk. Term structure models based on Equation (6.1) imply stochastic discount factors of the following form:

$$
\begin{aligned}
{}_t m_{t+1} &= \exp(-A_{t+1} - B_{t+1} - C_{t+1}), \quad \text{where} \\
A_{t+1} &= {}_t\int^{t+1} r_s \mathrm{d}s, \\
B_{t+1} &= {}_t\int^{t+1} q(X_s)\mathrm{d}w_s \\
C_{t+1} &= (1/2)_t\int^{t+1} q(X_s)^2 \mathrm{d}s,
\end{aligned}
\tag{6.2}
$$

where r_s is the instantaneous interest rate at time s. The notation ${}_t m_{t+1}$ is chosen to emphasize that the SDF refers to a discrete time interval, say one month, that begins at time t and ends at time $t + 1$. When there are multiple state variables there is a term like B_{t+1} and C_{t+1} for each state variable. A particular term structure model specifies the state variables, X, and the functions $\mu(\cdot)$,$\sigma(\cdot)$, and $q(\cdot)$.

One advantage of a continuous time model is that the SDF should correctly price dynamic strategies that trade as functions of the state variables, X_t. In particular, the monthly returns of mutual funds that result from interim trading can be handled without concern. Thus, the interim trading biases discussed above should not be a problem in this class of models.

Ferson et al. (2006a) use a first-order Euler approximation to Equation (6.1) in order to implement the models:

$$X(t + \Delta) - X(t) \approx \mu(X_t)\Delta + \sigma(X_t)[w(t + \Delta) - w(t)]. \tag{6.3}$$

The period between t and $t + 1$ is divided into $1/\Delta$ increments of length Δ. The period is one month, to match the mutual fund returns, divided into increments of one day. For a given model, we can use daily data on $X(t + \Delta)$ and $X(t)$ and the functions $\mu(X_t)$ and $\sigma(X_t)$ are specified. The approximate daily values of $[w(t + \Delta) - w(t)]$ are inferred from Equation (6.3). The terms A_{t+1}, B_{t+1} and C_{t+1} in Equation (6.2) are approximated using daily data by

$$A_{t+1} \approx \sum_{i=1,\ldots,1/\Delta} r(t+(i-1)\Delta)\Delta$$
$$B_{t+1} \approx \sum_{i=1,\ldots,1/\Delta} q[X(t+(i-1)\Delta)][w(t+i\Delta) - w(t+(i-1)\Delta)]$$
$$C_{t+1} \approx (1/2) \sum_{i=1,\ldots,1/\Delta} q[X(t+(i-1)\Delta)]^2\Delta. \quad (6.4)$$

Farnsworth (1997) and Stanton (1997) evaluate the accuracy of these first order approximation schemes. Stanton concludes that with daily data, the approximations are almost indistinguishable from the true functions over a wide range of values, and the approximation errors should be small when the series being studied is observed monthly. He also evaluates higher-order approximation schemes, and finds that with daily data they offer negligible improvements.

Ferson et al. (2006a) show that a popular class of three-factor affine term structure models, include the models of Cox et al. (1985) and Vasicek (1977) as special cases, imply the following reduced form expression for the approximated SDF:

$$\begin{aligned} {}_t m_{t+1} = \exp(&a + bA^r_{t+1} + c[r_{t+1} - r_t] + dA^l_{t+1} \\ &+ e[l_{t+1} - l_t] + f[c_{t+1} - c_t] + gA^c_{t+1}), \end{aligned} \quad (6.5)$$

where

$$A^r_{t+1} = \sum_{i=1,\ldots,1/\Delta} r(t+(i-1)\Delta)\Delta,$$
$$A^l_{t+1} = \sum_{i=1,\ldots,1/\Delta} l(t+(i-1)\Delta)\Delta,$$
$$A^c_{t+1} = \sum_{i=1,\ldots,1/\Delta} c(t+(i-1)\Delta)\Delta.$$

The coefficients $\{a, b, c \ldots\}$ are constant functions of the underlying fixed parameters of the models. The two-factor affine model is nested in the general three factor model of by setting $f = g = 0$. The single-factor affine model is nested in the two-factor affine model by further setting $d = e = 0$.

Note that the single-factor model actually depends on two short rate "factors." Because of the effects of time aggregation, there is both a discrete change in the short rate, $[r_{t+1} - r_t]$, and an average of the daily short-rate levels over the month. The two-factor affine model depends on the monthly changes in the long and short rates and on monthly averages of the long rate and short rate levels. The three-factor model adds a discrete change in convexity and an average convexity factor.

The time averaged terms in (6.5) are needed to control the interim trading bias. Thus the OE portfolio needs to include maximum correlation portfolios for time-averaged factors, in order to correct for interim trading biases. In practice we are limited by the data, and with daily data it is implicitly assumed that managers trade only at the end of each day. Funds actually engage in intradaily trading, of course, so the accuracy of this approach, using available data, remains an open empirical question.

7

Hedge Fund Performance

Hedge funds have been in business for over 60 years. However, the recent growth in hedge fund assets and the significant attention devoted to hedge funds in the popular press has increased the interest in hedge fund performance. Like mutual funds, hedge funds are open-ended investment companies that pool dollars from a group of investors. Unlike mutual funds, hedge funds are exempt from the Investment Company Act. The growth in the industry and the differences from mutual funds ~~has~~ created a challenging new avenue for performance evaluation.

From a research standpoint, hedge fund performance poses several interesting features. For one, the compensation of a typical hedge fund manager includes a significant performance fee. That is, a share of the fund's returns, often after the fund performance exceeds a "high water mark." These compensation contracts are thought to attract the most highly skilled managers. Therefore, the hedge fund industry should provide a fertile laboratory for detecting superior investment performance. Also, the secretive nature of the industry makes historical data limited and, when available, subject to a number of potential biases.

In addition, their broad investment mandates allow hedge funds the flexibility to trade frequently, thereby creating time-variation in risk exposures. A related issue, due to the lack of transparency about hedge fund strategy, is the problem of choosing an appropriate benchmark for measuring a fund's performance. Asset illiquidity also makes the reported returns challenging to evaluate. When asset values are non-synchronously measured, there may be biases in the measured performance. These distinguishing characteristics of hedge funds create interesting research challenges.

7.1 Data Issues

Academic research on hedge funds has mainly relied on five alternative commercial databases. These are the US Offshore Funds Directory, Managed Account Reports (MAR), the Hedge Fund Research (HFR) database, TASS and CSFB/Tremont. These databases have a central feature in common with the available databases on pension funds: they are self-reported on a voluntary basis by the funds to a data vendor. While the vendors strive to minimize errors, voluntary self-reporting creates several potential biases. These include survivorship biases, back-filling biases, and smoothing biases.

7.1.1 Survivorship Bias

Funds may leave a database because they performed poorly and are liquidated. If these "dead" funds are subsequently removed from the database, it introduces a form of survivorship bias in the reported returns. For example, the average returns of the surviving funds overstates the expected returns to fund investors. Funds may also choose to cease reporting to a database if they close to new investors and therefore do not place as much value on the "advertising" that such reporting implies. If successful funds die because they close to new investors with a record of success, the average bias works in the other direction. Since about 1993, hedge fund data providers like HFR have retained data for dead funds, which makes it possible to estimate survivorship biases, as discussed below.

7.1.2 Backfilling

Backfilling occurs when the database includes returns that preceded a fund's entry into the database. Backfilled returns may impart an upward bias in the estimated performance level due to fund incubation, whereby a manager of many funds chooses to continue only those funds that have above-average performance histories. Evans (2003) finds that mutual funds are "incubated" before they are opened to the public and enter the database, as described above. A similar effect occurs for pension funds and hedge funds. If funds choose to advertise by entering the database only when they have good track records, the initial reported returns will be biased upwards. Estimates of backfilling are limited because the databases often did not provide the dates the funds were added to the database. Park (1995) was among the first to identify hedge funds that are added to the databases with their prior history.

7.1.3 Smoothing

Since the reporting of returns in hedge fund databases is voluntary and largely unregulated, funds have some flexibility in what they report. Illiquid assets in particular raise the issue of smoothing in the reported returns. While mutual funds are restricted to hold no more than 5% of their portfolio in illiquid assets, hedge funds face no such restrictions. Illiquid assets may be carried at historical cost for some time, at the discretion of the fund, even after the true market value may have changed. This creates the opportunity to smooth reported returns, and thus present the impression of lower return volatility. Smoothing also creates biases in measures of systematic risk for hedge funds (Asness et al., 2001). Getmansky et al. (2004) show how infrequent valuation of hedge fund assets can induce autocorrelation, in the form of moving average components, into hedge fund returns. We discuss this in more detail below.

7.2 Dynamic Risk Exposures

We have discussed dynamic risk exposures in the context of conditional performance evaluation models, applied to mutual funds and pension funds. Dynamic trading strategies and risk exposures may to be especially relevant for hedge funds. One of the most widely recognized studies of the dynamic strategies of hedge funds is Fung and Hsieh (1997). Fung and Hsieh perform a style analysis using monthly returns of mutual funds and hedge funds. They estimate the extent to which fund return variation can be explained by traditional benchmarks, such as buy-and-hold returns from equities, bonds, commodities, and currencies. They find a striking difference in the distribution of R^2 between the two sample of funds. Specifically, 47% of mutual funds have R^2's above 75%, and 92% have R^2's above 50%. In contrast, nearly half (48%) of the hedge fund R^2's are below 25%. In addition, unlike mutual funds, a substantial fraction (25%) of hedge funds have negative exposure to the benchmarks. Fung and Hsieh explain the above results by noting that within an asset class, hedge funds can display much different trading activities and hence returns than mutual funds. Fung and Hsieh find that five mutually orthogonal principal components can explain on average 43% of the return variance in a sample of hedge funds. Thus, there is a lot of common variation in hedge fund returns. Interestingly, the returns on these factors are related to traditional benchmarks in a nonlinear way.

7.3 Asset Illiquidity

We mentioned above how illiquid assets can lead to smoothing of hedge fund returns data, but the liquidity is more than a data issue. Illiquid assets may earn a premium, in the form of a higher expected return. As hedge funds often hold illiquid positions, this raises the interesting economic question: Who gets the liquidity rents; the fund managers or the funds' investors? Aragon (2007) argues that the liquidity rents are split. Investors in many hedge funds are subject to "lock-up" periods, during which they are not allowed to withdraw their money. They often have to give the fund advanced notice if they intend to withdraw money. Aragon argues that investors earn higher returns associated with these restrictions. These restrictions, in turn, make it easier for hedge fund managers to pursue illiquid strategies with higher payoffs.

8

Recent Empirical Evidence

8.1 Evidence on Conditional Alphas

Ferson and Schadt (1996) find evidence that funds' risk exposures change significantly in response to variables that represent public information on the economy, such as the levels of interest rates and dividend yields. Using conditional models Ferson and Schadt (1996) and Kryzanowski et al. (1997) find that the distribution of mutual fund alphas is shifted to the right, relative to the unconditional alphas, and is centered near zero. Thus, conditional models tend to paint a more optimistic picture of mutual fund performance than unconditional models. This general pattern is confirmed in subsequent studies by Zheng (1999), Ferson and Qian (2004), and others. A zero alpha suggests that funds have enough investment ability to cover their fees and trading costs, but that they do not add value for investors. This is consistent with the model of Berk and Green (2004) in which managers are able to extract the full rents accruing to their investment ability.

Ferson and Warther (1996) attribute differences between unconditional and conditional alphas to predictable flows of public money into funds. Inflows are correlated with reduced market exposure, at times when the public expects high returns, which leads to larger cash holdings at such times. Larger cash holdings when returns are predictably higher leads to lower unconditional performance, but does not affect the CPE measures. In pension funds, which are not subject to high frequency flows of public money, no overall shift in the distribution of fund alphas is found when moving to conditional models (Christopherson et al., 1998).

8.2 Conditional Market Timing

While studies of mutual funds' market timing ability using the classical models found evidence of perverse, negative timing coefficients, the literature on conditional market timing finds different results. Ferson and Schadt (1996) showed that the classical measures can produce negative coefficients for naïve dynamic strategies. The conditional timing measures avoid this bias and suggested neutral timing ability in their sample of equity style mutual funds, 1968–1990.

Becker et al. (1999) simultaneously estimate parameters that describe the public information environment, the risk aversion of the fund manager and the precision of the fund's market-timing signal. Using a sample of more than 400 US mutual funds from 1976 through 1994, they find that both benchmark investing and conditioning information are important in the model. The point estimates suggest that mutual funds behave as highly risk-averse benchmark investors, but the standard errors of the risk-aversion estimates are large. A cross-sectional analysis of mutual fund holdings suggests that the model parameters are informative about managers' portfolio strategies. Once the public information variables are controlled for, there is little evidence that mutual funds have conditional market-timing ability. Only in a subsample of asset-allocation style funds is there a hint of timing ability, in the form of significantly-precise market timing signals.

Ferson and Qian (2004) allow for conditional timing coefficients to vary over time. They find that successful market timing is more likely when the stock market and short term corporate debt markets are highly liquid. This makes sense, as market timing trades may be made at lower cost in highly liquid markets. They also find that market timing funds can deliver significant conditional timing performance when the term structure slope is steep. In contrast, the timing funds seem unable to deliver reliable market timing services when the slope of the term structure is flat. This may reflect a larger dispersion in the cross-section of asset returns, and therefore more room for asset allocation strategies, when the term structure is steeply sloped. When they sort the timing funds into groups according to fund characteristics, the best conditional timers have the longest track records, the largest total net assets, the lowest expense ratios, or the smallest capital gains.

Busse (1999) asks whether fund returns contain information about market volatility. He finds evidence using daily data that funds may shift their market exposures in response to changes in second moments.

8.3 Evidence from Weight-Based Measures

Grinblatt and Titman (1993) are the first to apply weight-based performance measures to study the quarterly holdings of mutual funds. They use a sample of 155 funds over the period 1974–1984.[1] They compare the results obtained with Cornell (1979) measure with those obtained from a measure based upon changes in portfolio weights. Grinblatt and Titman take the fund's OE portfolio in a given quarter to be the fund's portfolio, as defined by its weights in the previous quarter. Thus, the underlying model assumes that a manager with no information holds fixed portfolio weights. Fund performance is measured as the average difference in raw returns over the subsequent quarter, between the fund and this OE portfolio. They estimate abnormal before cost performance to be a statistically significant 2% per year in aggregate. Aggressive growth funds exhibit the strongest performance at 3.4%. Since neither the funds' hypothetical returns not the OE portfolio returns pay any costs or management fees, the weight-based measure speaks directly to investment ability as we have defined it. Thus, the evidence suggests that fund managers in aggregate do have investment ability.

Daniel et al. (1997) study the holdings of most equity mutual funds that existed during any quarter within the period 1975–1994. For each fund in a given quarter, they define the OE portfolio based upon the

[1] Grinblatt and Titman (1989a) are the first to study the reported holdings of mutual funds. However, their focus is to compare the hypothetical returns constructed from holdings with the actual returns reported by the funds.

characteristics of the securities held. Specifically, each security in the fund's portfolio is assigned to one of 125 characteristic groups, depending upon its size, book-to-market ratio, and lagged return, measurable with respect to the beginning of the quarter. They construct passive (value-weighted) portfolios across all NYSE, AMEX, and NASDAQ stocks for each characteristic group. The return on the OE portfolio in a given quarter is the summation, across all securities in fund's portfolio, of fund portfolio weights times the return on the corresponding passive portfolio. They find that the average fund in their sample has a significantly positive performance of 77 basis points per year. This is further evidence of investment ability. Daniel et al. (1997) conclude that the observed performance is about the same size as the typical management fee. Therefore, the average fund is not expected to deliver value added to investors after costs.

Ferson and Khang (2002) examine the conditional, weight-based approach to measuring performance. They first conduct some experiments to assess the extent of interim trading bias in returns-based measures. Even though their hypothetical portfolios trade only two times between each quarterly observation date, they find that the interim trading bias has a huge impact on returns-based performance measures. The average Jensen's alpha of the strategies, computed relative to the value-weighted CRSP market, is 1.03% per quarter with an average t-ratio of 3.64. Thus, by mechanically trading with public information, it is possible to generate large and economically significant alphas.

Ferson and Khang apply weight-based measures to a sample of pension fund managers, over the 1985–1994 period. They find that under the unconditional weight measures growth managers show small positive performance. Using Conditional Weight Measures to avoid interim trading bias and to control for public information effects, the estimates of the pension funds' performance are close to zero.

8.4 Stochastic Discount Factor Evidence

Chen and Knez (1996) were the first to use conditional SDF alphas to evaluate performance in a sample of US equity style mutual funds. They used an SDF formed from a set of primitive assets such that the OE portfolio is mean–variance efficient in those assets. With this measure, they find that mutual funds have insignificant but negative abnormal returns. With unconditional models, they find the average alpha for 68 funds, 1968–1989, is −0.09% per month.

Farnsworth et al. (2002) use a variety of SDF models to evaluate performance in a monthly sample of US equity mutual funds. They find that many of the SDF models are biased. The average bias is about −0.19% per month for unconditional models and −0.12% for conditional models. This is less than two standard errors, as a typical standard error is 0.1% per month. Farnsworth et al. conclude that the findings of Chen and Knez (1996) reflect a biased performance measure.

Dahlquist and Soderlind (1999) study SDF models similar to those of Chen and Knez, using weekly Swedish data, 1986–1995. They find no significant biases in the average pricing errors, but they do find size distortions, where tests for the hypothesis that $\alpha_p = 0$ reject the null hypothesis too often. Using simulations Farnsworth et al. (2002) find that the average mutual fund SDF alpha is no worse than a hypothetical stock-picking fund with neutral performance. Adding back average expenses of about 0.17% per month to the mutual fund alphas, the average fund's performance is slightly higher than hypothetical funds with no investment ability. Thus, the overall evidence on equity fund performance using conditional SDF methods is similar to the evidence using conditional beta pricing models.

8.5 Pension Fund Evidence

Coggin et al. (1993) study the market timing and stock picking ability of US equity style pension funds and provide references to the few academic studies of pension fund equity manager performance that were available at the time. Lakonishok et al. (1992) find that the typical pension fund earns average returns about 1% below that of the SP500 index over 1983–1989, which they interpret as evidence of poor performance. However, Christopherson et al. (1998a) show that pension funds tend to hold smaller stocks than those of the SP500 on average. They also find that Lakonishok et al. evaluated their funds during a sample period when small stocks returned less than large stocks. Using a more appropriate style-based benchmark that controls for the market capitalization of the stocks, Christopherson et al. found evidence for positive pension fund alphas.

Christopherson et al. (1998a) study the conditional performance of equity style US pension funds. They find evidence consistent with selection bias for the average returns in the pension fund sample. The average annual return on an equally weighted portfolio of all managers is 16.11% over 1981–1990. Excluding the first five years of data for each manager, the average drops to 15.45%. In their analysis of performance persistence they use the returns following the first five years of data for a given manager. They find persistent performance, concentrated in the managers with poor prior-period performance measures. A conditional approach, using time-varying measures of risk and abnormal performance, is better able to better detect this persistence and to predict the future performance of the funds than are traditional methods.

In their study of equity pension fund managers, 1985–1994, Ferson and Khang (2002) find that the traditional, returns-based alphas of the funds are positive, consistent with previous studies of pension fund performance. However, these alphas are smaller than the potential effects of interim trading bias. By using instruments for public information combined with portfolio weights, their conditional weight-based measures find that the pension funds also have neutral performance. Thus, the overall empirical evidence based on conditional performance measures suggests that abnormal fund performance of equity style pension funds, controlling for public information, is rare.

8.6 Evidence on Bond Fund Performance

In 2003, the total net assets of US bond funds exceeded 1.2 trillion dollars, about 1/6 the amount in equity-style mutual funds and similar to the value of hedge funds.[2] Large amounts of fixed-income fund assets are also held in professionally managed portfolios outside of mutual funds, for example in pension funds, trusts and insurance company accounts. Thus, it is important to understand how to evaluate the

[2] These figures are from the Investment Company Institute.

performance of bond fund managers. However, at this writing the number of academic studies on bond funds is relatively small.

Elton et al. (1993, 1995) and Ferson et al. (2006a,b) study US bond mutual fund performance, concentrating on the funds' risk-adjusted returns, or alphas. They find that the typical average performance after costs is negative and on the same order of magnitude as the funds' expenses. They also find that the effects of controls for interim trading bias in the measured performance of bond funds is important. Brown and Marshall (2001) develop an active style model and an attribution model for fixed income funds, isolating managers' bets on interest rates and spreads. Comer et al. (2005) study timing ability in a sample of 84 high quality corporate bond funds, 1994–2003, using variations on Sharpe's (1992) style model. Comer (2006) measures bond fund sector timing using portfolio weights. Aragon (2005) studies the timing ability of balanced funds for bond and stock indexes.

Chen et al. (2006) study the ability of US bond funds to time factors related to bond markets, controlling for non-timing-related sources of nonlinearity in the relation between fund returns and the factors. Nonlinearities appear in the underlying assets held by funds and may also arise from dynamic trading strategies or derivatives, funds' responses to public information, or systematic patterns in stale pricing. Chen et al. find that controlling for non-timing-related nonlinearity is important. Bond funds' returns are typically more concave than passive benchmark returns, relative to nine common factors, without controls for the non-timing-related nonlinearities. This would appear as poor timing ability in naive models. With the controls, the distribution of the timing coefficients appears neutral at the fund style-group level. A cross-sectional analysis at the individual fund level finds that the measured timing coefficients have little explanatory power for future returns. Overall, they find no evidence that bond funds have timing ability.

8.7 Evidence on Hedge Fund Performance

8.7.1 Early Evidence

Brown et al. (1999) are among the first to provide a comprehensive examination of hedge fund performance. They collect a sample of funds from the US Offshore Funds Directory and study the annual net-of-fees annual returns over the 1989–1995 period. Jensen's alpha is 16.6% and 5.7% per year for equal and value-weighted portfolios of the hedge funds, respectively. They interpret these results as evidence of superior manager skill. However, Brown et al. (1999) note that their evidence could potentially be explained by a survivorship bias, because they cannot track the performance of funds that disappear within the year.

Ackermann et al. (1999) study monthly hedge fund returns over the 1988–1995 period. The data are taken from Managed Account Reports (MAR) and Hedge Fund Research (HFR). Their use of monthly data reduces the survivorship bias relative to Brown et al. (1999). Ackermann et al. (1999) calculate the average alphas of equal-weighted hedge fund style portfolios, relative to the SP500 index, and assuming that the funds betas are all equal to 1.0. They find excess returns average an insignificant −0.40%. However, after controlling for the actual betas of hedge funds (a median value of only 0.28 ~~relative to the SP 500~~), they find a significant positive Jensen's alpha for the equal-weighted hedge fund portfolio.

Ackermann et al. (1999) also examine the relation between individual hedge fund Sharpe Ratios and fund characteristics. They find that a fund's performance fee is statistically positively related to its Sharpe Ratio. For example, an increase in a fund's performance fee from zero to the median (20%) corresponds to a Sharpe Ratio that is larger by 0.15. They conclude that the performance fee aligns the interests of managers and investors, and that a higher incentive fee makes managers work harder to achieve higher average returns.

Liang (1999) studies monthly returns of 921 hedge funds from Hedge Fund Research, Inc. He compares the performance of 16 equal-weighted hedge fund style indices. Liang estimates risk-adjusted performance as the intercept in a regression of monthly returns for each hedge fund index on traditional benchmarks such as equity, bond, commodities, and currencies. He restricts his analysis to 385 funds with consecutive monthly returns from January 1994 to December 1996. He finds positive alphas for 11 of the 16 style groups, and 7 are statistically different from zero. He interprets these results as evidence of superior manager skill.

Liang (1999) also examines the relation between average returns and fund characteristics. Consistent with Ackermann et al. (1999), he finds that average returns are significantly related to the level of a fund's performance fee. Specifically, average returns are higher for funds with higher performance fees, for younger funds, larger funds, and funds that restrict investors through a lockup.

8.7.2 Evidence on Survivorship Bias

The voluntary nature of hedge fund reporting creates potential biases due to survivor selection, as discussed above. The sample in Brown et al. (1999) does not include funds that might have disappeared from the database during the year. They compare the performance of the equal-weighted index of all funds with an equal-weighted index of all funds that are present at the end of the sample period. The latter index tracks the performance of funds that do not leave the database during the seven-year sample period. Brown et al. (1999) obtain an estimate of survivorship bias by taking the difference in performance between these two indices. They report that the survivorship bias can be as large as 3% per year.

Fung and Hsieh (2000) examine hedge fund returns from the TASS database over the 1994–1998 period. They compare the average returns of an equal-weighted portfolio of all funds with an equal-weighted portfolio of surviving funds. They estimate a survivorship bias of 3%, similar to the figure of Brown et al. (1999).

Liang (2000) investigates survivorship bias by comparing the TASS and HFR databases. He calculates the performance difference between the sample of surviving funds only and the complete sample of funds. He finds a survivorship bias of 0.60% per year for the HFR database and 2.24% per year for the TASS database. Liang attributes the difference in bias to the fact that TASS contains a larger proportion of offshore funds. He argues that, compared to an onshore fund, an offshore fund that disappears from the database is more likely to have had poor performance.

Overall, the evidence suggests that survivorship bias is a potentially significant issue for the evaluation of hedge fund performance.

While the bias could in principle be positive or negative, the estimates suggest a positive bias is likely. Thus, the measured performance in a sample of funds that survive until the end of the measurement period is probably an overly optimistic indicator of what can be expected in the future.

8.7.3 Evidence on Backfilling

Park (1995) was among the first to study funds are added to the databases with their prior history. He examines a sample of Commodity Trading Advisors (CTA's) from the MAR database, and reports an average track record length of 27 months at the time the fund is added to the database. Ackermann et al. (1999) provide an indirect estimate of backfilling bias by comparing the average returns of hedge funds with and without each fund's first two years of performance data. They find that the statistics for the two samples of funds closely track each other, and conclude that there is no backfilling bias. However, the Ackermann et al. (1999) data cover the period January 1988–December 1995, and the amount of hedge fund data mushrooms after 1993. Backfilling bias is likely to be very different in the post-1993 data.

Fung and Hsieh (2000) find an average incubation period of 343 days for hedge funds during 1994–1998. They estimate backfilling bias by comparing equal weighted portfolio returns using the complete sample with equal weighted portfolio returns using the sample after dropping the initial 12 months of returns for each fund. The difference is 1.4% per year.

Posthuma and van der Sluis (2003) study the TASS database over 1994–2002 and estimate backfilling using the dates the funds were added to the database. They first find that the mean and median number of backfilling months is 34 and 23 months, respectively. These numbers are much larger than the mean backfilling time estimated by Brown et al. (1999) (15 months) and Fung and Hsieh (2000) (343 days). They estimate backfilling bias by calculating the annual difference between the backfilled and non-backfilled index. Their figure is 4.35% per year.

Overall, the evidence suggests that backfilled return histories are overly optimistic indicators of subsequent hedge fund performance.

The best practice is probably to avoid measuring returns before the date that the fund opened its doors to investors, when this is possible.

8.7.4 Evidence on Hedge Fund Strategies

Fung and Hsieh (2001) study a sample of 1304 CTA's over the 1989–1997 period and find that the funds' returns resemble a straddle on a broad equity index. In particular, fund returns are positive whenever the market index realizes exaggerated positive or negative returns. Fung and Hsieh (2001) use option price data to construct the returns of option straddles on 26 different markets, including stock, bond, currency, and commodity markets. They find that the first principal component from a factor analysis of the funds has little exposure to systematic risk relative to a linear factor model using standard benchmarks. For example, regressing the trend-following returns against the eight major asset classes results in an adjusted-R^2 of just 1%. In addition, none of the factor betas are statistically significant. In contrast, a regression of the hedge fund portfolio returns against the straddle returns results in an adjusted-R^2 of 47.9%. They conclude that trend following hedge funds do have systematic risks, but that the dynamic trading strategies lead to nonlinearity, and this leads traditional performance models to conclude that funds have much lower systematic risk.

Agarwal and Naik (2004) analyze the returns of equity-oriented hedge fund indexes over the 1990–2001 period. They augment standard linear factor models with the monthly returns of at-the-money and out-of-the-money European call and put options on the Standard and Poors Composite Index. They find that a large number of the indexes have significant exposure to the options. For example, the Event Arbitrage style index has a significant negative return beta of −0.94, measured against out-of-the-money put options. This makes sense, as Mitchell and Pulvino (2001) show that the gains from a successful merger do not depend on the market index, except in severely depreciating markets. Agarwal and Naik (2004) conclude that option-based factors are an important control for hedge fund performance studies.

8.7.5 Evidence on Asset illiquidity

Asness et al. (2001) were the first to study the implications of hedge fund illiquidity on estimates of performance. They estimate Jensen's alpha regressions like Equation (2.2) using returns on the CSFB/Tremont hedge fund indices over the 1994–2000 period. Relative to the SP500 index, hedge fund returns produce an average unconditional beta of 0.37 and an average alpha of 2.63% per year. Asness et al. (2001) compare these results with those obtained from the modified regression, proposed by Scholes and Williams (1977) and Dimson (1979), that includes both contemporaneous and lagged market returns on the right-hand side in order to account for non-synchronous trading. The idea is that the covariance of the current measured return of the fund with the lagged index return should capture "stale" measured returns due to illiquid assets that did not trade in the current period. The Scholes–Williams beta estimate is the sum of the slope coefficients on the current and lagged index returns. For the Aggregate Hedge Fund Index, Asness et al. (2001) estimate the summed beta to be 0.84, significantly larger than the 0.37 produced with the standard market model. The annualized alpha becomes negative and statistically insignificant after including lagged market returns. They also find that the importance of lagged beta coefficients for hedge fund returns is greater for categories that are commonly viewed as illiquid trading styles. They conclude that illiquidity is an important consideration for hedge fund performance studies.

Getmansky et al. (2004) also examine the role of asset illiquidity on the performance measurement of hedge funds. They develop a returns-based proxy for fund asset illiquidity, based on the autocorrelation that stale prices create in measured returns. They estimate their liquidity measure for a sample of 908 hedge funds from the TASS database. They find lower liquidity measures in styles that are commonly viewed as involving illiquid assets.

Using monthly data from 1994 to 2001, Aragon (2007) finds a positive, concave relation between hedge funds' after-fee excess returns and share restrictions, including lockup restrictions, redemption notice periods, and minimum investment amounts. The difference in the excess returns between funds with and without lockup restrictions, or the lockup premium, is 4%–7% per annum. Strikingly, after controlling for lockup, notice period length, and minimum investment size, the previously positive alphas are either negative or insignificant. Aragon (2007) also finds a relation between a fund's asset illiquidity, as measured as in Getmansky et al. (2004), and the decision to impose share restrictions. He concludes that restrictions enable funds to earn illiquidity rents, because they reduce costly trading by the funds' clients, and this allows a fund to hold assets with greater illiquidity premiums.[3] Investors can expect higher returns on funds with share restrictions, commensurate with the illiquidity they bear in such funds. ~~Aragon argues that hedge fund returns are consistent with the required returns for holding illiquid fund shares.~~

[3] This is supported by Edelen's 1999 finding that unrestricted mutual fund investor trading can reduce fund performance by 1%–2% per year. Chordia (1996) argues that the adverse effects of investor flows in mutual funds are greater for funds holding more illiquid assets.

9

A Summary: The Evidence on Managed Portfolio Performance and Market Efficiency

The evidence on the performance of professionally managed portfolios relates to the classical question of the informational efficiency of the markets, as summarized by Fama (1970). This section first describes how these ideas are related and then presents some tables that summarize the empirical evidence.

9.1 Market Efficiency and Portfolio Performance

As emphasized by Fama (1970), any analysis of market efficiency involves a "joint hypothesis." There must be an hypothesis about the model for equilibrium expected returns and also an hypothesis about the informational efficiency of the markets. These can be described using the representation for asset pricing models in Equation (4.1). Assume that Equation (4.2) holds when the conditioning information is Ω_t:

$$E\{m_{t+1}R_{t+1}|\Omega_t\} = 1, \tag{9.1}$$

where Ω_t refers to the information that is conditioned on by agents in the model, and in that sense "reflected in" equilibrium asset prices.[1] The hypothesis about the model of market equilibrium in the joint hypothesis amounts to a specification for the stochastic discount factor, m_{t+1}. For example, the CAPM of Sharpe (1964) implies that m_{t+1} is a linear function of the market portfolio return (e.g., Dybvig and Ingersoll, 1982), while multibeta asset pricing models imply that m_{t+1} is a linear function of the multiple risk factors.

Equation (4.3) defines the SDF alpha, where we use the law of iterated expectations to replace Ω_t with the observable instruments, Z_t. Note that if the SDF prices a set of "primitive" assets, R_{t+1}, then according to Equation (9.1) α_{pt} will be zero when a fund (costlessly) forms a portfolio of the primitive assets, if the portfolio strategy uses only information contained in Ω at time t. In that case $R_{p,t+1} = x(\Omega_t)'R_{t+1}$, where $x(\Omega_t)$ is the portfolio weight vector. Then $\alpha_{pt} = E\{[E(m_{t+1}x(\Omega_t)'R_{t+1}|\Omega_t)] - 1|Z_t\} = E\{x(\Omega_t)'[E(m_{t+1}R_{t+1}|\Omega_t)] - 1|Z_t\} = E\{x(\Omega_t)'\underline{1} - 1|Z_t\} = 0$. Informational efficiency of the market says that you cannot get a conditional alpha different from zero using any information that is contained in Ω_t.

Fama describes increasingly fine information sets in connection with market efficiency. Weak-form efficiency uses the information in past stock prices to form portfolios of the assets. Semi-strong form efficiency uses variables that are obviously publicly available, and strong form uses anything else. The different information sets described by Fama (1970) amount to different assumptions about what information is contained in Ω_t. For example, weak-form efficiency says that past stock prices cannot be used to generate alpha while semi-strong form efficiency says that other publicly available variables would not generate alpha.

In summary, informational efficiency says that you cannot get an alpha different from zero using any information Z_t that is contained in Ω_t. Since alpha depends on the model through m_{t+1}, there is always

[1] If X_{t+1} is the payoff and P_t is the price, then $R_{t+1} = X_{t+1}/P_t$ and Equation $(\cdot)$ says that $P_t = E\{m_{t+1}X_{t+1}|\Omega_t\}$. The equilibrium price is the mathematical conditional expectation of the payoff given Ω_t, "discounted" using m_{t+1}. In the language of Fama (1970), this says that the price fully reflects Ω_t.

a joint hypothesis at play. Indeed, any evidence in the literature on market efficiency can be described in terms of the joint hypothesis; that is, the choice of m_{t+1} and the choice of the information Z_t.

How does the evidence on the performance of professionally managed portfolios relate to informational efficiency? All of the fund performance evidence can be described as examples of this simple framework. However, two complications arise with examples of fund performance. One is the issue of costs and the other relates to who is using the relevant information. With respect to costs, we use *investment ability* versus *value added* to distinguish performance on a before-cost versus after-cost basis. Studies of market efficiency also consider trading costs, and serious violations of efficiency are usually considered to be those that are observed on an after-cost basis. Thus, our concept of value added is closely related to market efficiency.[2] If we find that a manager has value added in a conditional model that controls for public information, this rejects a version of the joint hypothesis of semi-strong form efficiency. If we do not question the model for m_{t+1} (and the associated OE benchmark) then we may interpret such evidence as a rejection of the informational efficiency part of the joint hypothesis.

The second complication relates to whether the portfolio manager or other investors are using the information in question. We have described efficiency in terms of the information in portfolio weights. At the fund level, managers use their information to form the fund's portfolio weights. Evidence about the performance of a fund therefore relates to the information used by the manager. However, much of the evidence in the literature on fund performance is described in terms of portfolio strategies that combine mutual funds. For example, a manager may use private information to deliver alpha, which speaks to strong form efficiency. If these alphas persist over time and investors can use

[2] Grossman and Stiglitz (1980) point out that no one would expend resources to gather information if it did not pay to trade on it. So, it would be hard to imagine an efficient market if no one had investment ability. Investors only see mutual fund returns after the managers have been paid out of the fund's assets. The question of how fund managers are paid for their investment ability has to do with the efficiency of the labor market for fund managers. The value added for investors, on the other hand, is traditionally the central issue for studies of the efficiency of financial markets.

the information in the past returns of the funds to form strategies that deliver value added performance, this speaks to weak form efficiency.

9.2 Mutual Fund Examples

This section presents a summary of the evidence on mutual fund performance. In Table 9.1 we use monthly returns data on individual funds

Table 9.1 Summary of mutual fund performance evidence.

Fund style	Growth	Income	Sector	Small-firm	Timers
N	2069	1137	545	811	799
Avg. expense ratio	1.41	1.01	1.65	1.54	1.27
Sharpe Ratios					
Average	0.14	−0.24	0.13	0.13	0.15
% < Market	33.6	68.9	77.1	85.1	31.3
Jensen's Alphas					
Average (%)	−0.04	−0.05	0.21	−0.00	−0.10
Percent < 0	55.9	75.1	38.1	57.0	71.2
Avg. regression R^2	73.8	28.2	36.6	53.3	73.7
Style Alphas					
Average (%)	0.04	−0.06	0.36	0.11	0.04
Percent < 0	45.6	68.0	26.2	47.6	44.3
Conditional CAPM α					
Average (%)	0.01	−0.10	0.27	0.25	−0.11
Percent < 0	55.9	66.6	36.3	46.9	70.1
Avg. regression R^2	78.4	41.3	44.8	58.7	78.3
Unconditional Timing					
Avg. coefficient	−0.25	0.24	−1.23	−1.94	0.12
Percent < 0	54.8	54.4	69.0	91.8	38.5
Conditional Timing					
Avg. coefficient	−0.11	0.39	−1.24	−1.89	0.18
Percent < 0	50.6	54.4	67.7	89.8	36.5

Note: The sample starts in January of 1973 or later, depending on the fund group, and ends in December of 2000. The lagged conditioning variables for the conditional models are a 3-month Treasury yield, term slope, dividend/price ratio for the CRSP value-weighted stock market, a credit-related yield spread and a spread of 90-day commercial paper over treasury yields. *N* is the number of funds with at least 24 monthly returns. Alphas are reported in percent per month. The average expense ratios are shown, in annual percent. The average R-squares are shown for the regressions estimating alphas, in percent. The conditional CAPM alpha is estimated using the Ferson and Schadt (1996) model. The unconditional timing model is the Treynor–Mazuy regression, where the coefficient on the squared market excess return is summarized. The conditional timing model is the conditional Treynor–Mazuy model developed by Ferson and Schadt.

over the 1973–2000 period. The data are the same as in Ferson and Qian (2004). Funds are grouped into five styles: growth, income, sector, small company growth, and timers. The latter category includes balanced funds and asset allocation style funds, those types most likely to be attempting to time the markets. (See Ferson and Qian, 2004 for more detail.) The number of funds with at least 24 monthly returns ranges from 545 sector funds to 2069 growth funds (data for sector and small company growth funds start in 1990). The average annual expense ratios range from 1.01% for the income funds to 1.65% for the sector funds. As discussed by Ferson and Qian (2004), expense ratios have trended up during the sample period.

The first performance measure is the Sharpe ratio. Over this period the monthly Sharpe ratios vary between −0.24 for income funds and 0.15 for market timers. The ratio for the CRSP stock market over this period is 0.12. Many funds turn in lower Sharpe ratios than the stock index. More than one third of the growth funds and timers have lower Sharpe ratios than the index, while more than 85% of the small firm growth funds have lower ratios. Of course, the Sharpe ratio does not reflect portfolio diversification benefits that funds may offer, so we cannot conclude from a low ratio than investors would wish to shun these funds.

The next measure is Jensen's alpha. The averages range between −0.10% per month for timing funds to 0.21% for the sector funds. The results for the sector funds reflect only the 1990–2000 period, during the "dot com" boom. Over longer sample periods alphas tend to be negative on average, and we find negative average alphas for four of the five style groups. To interpret these figures, recall the joint hypothesis related to market efficiency. The model of market equilibrium assumes that the stochastic discount factor, m, is linear in the market index return, as in the CAPM. As a result, the OE portfolio is the market index adjusted with a fixed allocation to cash. These figures say that the mutual funds — with the possible exception of the sector funds — delivered no value added, or after cost return over this period, to an investor holding the market index and cash who ignores taxes and who pays negligible transactions costs for holding the market index and cash. The averages hide the fact that many funds have negative alphas.

More than 56% of the growth funds and more than 38% of the sector funds had negative alphas. Thus, an investor selecting even among the sector funds faces a significant risk of choosing a negative alpha fund. While the evidence in the literature on the persistence of funds' alphas is mixed, our reading of the literature suggests that much of the ability of past alpha to predict future alpha, resides in the negative alpha funds. So, it may be difficult for investors to capture the small value added that these figures might suggest, even for the sector funds.

If we add back the funds' average expense ratios to the alphas, the pre-expense ratio performance averages about 1.64% per year. Maintaining the market index as a good benchmark, we can interpret this figure as evidence of investment ability, on average, among the fund managers.

The table also reports style-based alphas, where the market index is replaced with a fund style-group specific benchmark. The style indexes are formed as in Ferson and Qian (2004), as portfolios of eight asset class returns that range from short term Treasuries, to corporate bonds, to portfolios of value and growth stocks. The weights used in the portfolio are estimated separately for each fund style group. (See Ferson and Qian for details.) The style alphas are less negative than the market-based alphas on average, and smaller fractions of the funds deliver negative style alphas. Sector funds and small firm growth funds, in particular, look better with a style benchmark. Part of this may reflect the poor relative performance of small stocks during the 1990s, so using a benchmark that puts more weight on small stocks and less weight on large stocks makes these funds look better.

The next measure is the conditional CAPM alpha of Ferson and Schadt (1996). We report the averages of the regression R-squares for the conditional model regression. These may be compared with the R-squares of the Jensen's alpha regression. The difference reflects the explanatory power of the interaction terms between the market index and the lagged conditioning variables, which captures time-variation in the conditional betas. The improvement in the R-squares is on the order of 5%–12%, which is evidence of time-varying fund betas. This is consistent with the findings in the literature, that mutual funds' betas tend to vary over time.

The conditional alphas in Table 9.1 are a little more optimistic about fund performance than the Jensen's alphas in many, but not all, of the cases. For the two fund groups measured over 1990–2000, the average is 26 basis points per month; for the three groups measured over 1973–2000, the average is -7 basis points per month. The fractions of individual funds with negative alphas range between 36% and 70%. Previous studies over different sample periods typically find negative Jensen's alphas and conditional alphas closer to zero, suggesting neutral value added on average. Our figures are broadly consistent with this. The interpretation of the conditional alphas is similar to that of Jensen's alpha, except now the OE portfolio combines the market index and cash with a time-varying weight that reflects a fund's time-varying beta.

The next two measures explore market timing ability. The unconditional measure is the coefficient on the squared market excess return in the Treynor–Mazuy (1966) quadratic timing regression. The average estimates are negative for three of the five style groups. Large fractions of the individual funds have negative timing coefficients: ranging from just over half of the income funds to more than 90% of the small company growth funds. This evidence is broadly consistent with much of the literature, which often interprets this as poor timing performance.

The next measure is a conditional version of the timing coefficient, based on Ferson and Schadt's generalization of the Treynor–Mazuy regression. Consistent with previous studies, the conditional timing coefficients present a slightly less negative impression about timing ability. Smaller fractions of the funds turn in negative conditional timing coefficients in each style group, and about 2/3 of the funds in the timing group record positive coefficients, but the differences are not great. Previous studies over different time periods also find slightly better results with respect to conditional timing coefficients (e.g., Becker et al., 1999), but it still seems puzzling to find so many funds that pursue market timing strategies without clear evidence of success with such strategies.

We do not present results using weight-based measures in Table 9.1. Weight-based measures of performance construct hypothetical returns using the funds reported portfolio holdings and returns data on the underlying securities. Thus, trading costs and expenses are not accounted for. The early measures typically produced returns larger than the returns of the benchmarks, the difference often being on the order of funds' expense ratios. This again suggests the presence of investment ability, but not value added. While the evidence is sparse using conditional weight-based measures, the conditional weight-based performance of pension funds, as measured by Ferson and Khang (2002), is close to zero. This suggests that the investment ability of pension funds can largely be captured through publicly available information. We think that more research is needed using conditional weight-based models to address this question for other kinds of funds.

The literature has found mixed evidence on the question of the persistence of fund performance. Persistence is the crucial issue for investors who wish to find high-return funds: Can a fund that performed relatively well in the past be expected to do so again in the future? It seems that certain characteristics of mutual funds, such as their levels of volatility and style choices, have some persistence over time. But, aside from effects such as momentum that can largely be explained by funds holdings of momentum stocks, the evidence suggests that good performance does not persist to any reliable degree. Perhaps, the best use of past relative performance information in mutual funds is to avoid persistently poor performers.

9.3 Hedge Fund Examples

The performance of hedge funds looked promising when academics studies first began to explore it empirically, as hedge funds delivered large alphas in traditional linear beta models. The unique incentive structures and other aspects of the industry suggested that the better managers may be found in this sector. However, as this literature has matured, it may be that the large alphas of hedge funds can be explained through a combination of data biases, such as survivor selection and backfilling, dynamic trading and nonlinear payoffs, asset illiquidity and infrequent trading.

This section presents a summary of the evidence on hedge fund performance. We use monthly returns data on individual funds over the

1994–2005. The data are provided by Lipper/TASS, which is one of the major hedge fund databases used in the literature. We consider both live funds and those funds that have disappeared from the database prior to December 2005. Net returns have already been reduced by the management and performance fees paid by investors to the fund manager, and by the costs of trading. Funds are grouped into one of eleven style categories based upon self-reported primary style categories. These include: Convertible Arbitrage (CA), Dedicated Short Bias (DSB), Emerging Markets (EM), Equity Market Neutral (EMN), Event Driven (ED), Fixed Income Arbitrage (FIA), Fund of Funds (FOF), Global Macro (GM), Long/Short Equity Hedge (LS), Managed Futures (MF), and Multi-Strategy (MS). As discussed in Section 7.1.2, funds typically bring a history of returns data with them upon being added to the databases, thereby creating a potential backfilling bias. We therefore restrict the analysis to nonbackfilled data.

Table 9.2 shows that the number of funds with at least 24 monthly returns ranges from 24 in the DSB group to 1062 for the LS group. The average percentage management and performance fees range from 1.2% to 2.3% and 9.1% to 19.5%, respectively. The proportion of funds that use a high watermark to calculate performance fees ranges from 28% (MF) to 66% (FIA).

The first performance measure is the Sharpe Ratio. Over this period the average monthly Sharpe ratios range from −0.03 for short sellers to 0.42 for Fixed Income Arbitrage funds. The majority of funds deliver Sharpe Ratios exceeding that of the market index for nearly all style categories. The lone exception is the Global Macro category, for which only 46% of funds beat the market Sharpe Ratio. Overall, this suggests that stock market investors could have improved their asset allocation by investing in the hedge fund industry during this period.

The next measure is Jensen's alpha. For all style categories, the nonbackfilled alpha is positive on average and also positive for the majority of funds. For example, the average monthly alpha is 0.11% for Global Macro funds and 0.48% for the Emerging Markets category. These results say that hedge funds appear to have delivered positive value added to an investor holding the market index and cash. This contrasts with the evidence of the previous section that mutual funds

Table 9.2 Summary of hedge fund performance evidence.

Fund style	CA	DSB	EM	EMN	ED	FIA	FOF	GM	LS	MF	MS
N	129	24	212	175	267	140	790	168	1062	369	115
Average Fees											
Management fee	1.3	1.2	1.5	1.3	1.3	1.2	1.5	1.5	1.2	2.3	1.4
Incentive fee	18.4	19.0	16.8	19.3	18.7	19.5	9.1	18.0	18.9	17.9	18.1
Highwater-mark	0.63	0.58	0.43	0.61	0.63	0.66	0.44	0.46	0.61	0.28	0.61
Sharpe Ratios											
Average	0.37	−0.03	0.23	0.17	0.34	0.42	0.26	0.08	0.18	0.05	0.31
% < Market	0.36	0.50	0.37	0.40	0.20	0.29	0.30	0.54	0.32	0.47	0.23
Jensen's Alphas											
Average (%)	0.28	0.28	0.48	0.19	0.43	0.24	0.18	0.11	0.43	0.21	0.47
Percent < 0	0.21	0.33	0.26	0.28	0.15	0.16	0.20	0.30	0.22	0.30	0.12
Avg. regression R^2	0.04	0.51	0.17	0.08	0.12	0.02	0.16	0.06	0.21	0.06	0.13
Style Alphas											
Average (%)	0.06	0.15	0.13	0.13	−0.07	0.16	−0.19	−0.41	−0.02	0.03	0.13
Percent < 0	0.33	0.42	0.36	0.30	0.44	0.20	0.56	0.53	0.40	0.33	0.25
Avg. regression R^2	0.39	0.54	0.38	0.02	0.29	0.14	0.45	0.11	0.25	0.33	0.16

(*Continued*)

Table 9.2 (*Continued*).

Fund Style	CA	DSB	EM	EMN	ED	FIA	FOF	GM	LS	MF	MS
Lagged Market Model Alphas:											
Average (%)	0.18	0.43	0.25	0.16	0.33	0.09	0.05	−0.08	0.30	0.18	0.38
Percent < 0	0.26	0.29	0.30	0.31	0.16	0.21	0.27	0.36	0.28	0.29	0.18
Avg. regression R^2	0.06	0.53	0.19	0.09	0.19	0.07	0.23	0.07	0.24	0.06	0.17
Unconditional Timing											
Avg. coefficient	0.09	0.21	−0.69	0.03	−0.18	−0.05	−0.19	−0.10	−0.25	0.34	−0.17
Percent < 0	0.27	0.38	0.66	0.45	0.68	0.44	0.64	0.42	0.60	0.18	0.49
Avg. regression R^2	0.07	0.51	0.18	0.09	0.14	0.04	0.17	0.06	0.22	0.08	0.14

Note: The sample starts in January 1994 or later, depending on the fund, and ends in December 2005. The estimation excludes all return observations that precede the data the fund was added to the database. Fund styles are Convertible Arbitrage (CA), Dedicated Short Bias (DSB), Emerging Markets (EM), Equity Market Neutral (EMN), Event Driven (ED), Fixed Income Arbitrage (FIA), Fund of Funds (FOF), Global Macro (GM), Long/Short Equity (LS), Managed Futures (MF), and Multi-Strategy (MS). N is the number of funds with at least 24 monthly returns. Alphas are reported in percent per month. The average management and incentive fees are shown, in annual percent. High watermark equals one if the fund uses a high watermark to calculate incentive fees. The average R-squares are shown for the regressions estimating alphas, in percent. The lagged market model is the Jensen's alpha regression including three lags of the market excess return. The timing model is the Henriksson–Merton regression, where the coefficient on the positive part of the market excess return is summarized.

deliver no value added to an investor who passively holds the market index and cash. However, our conclusions for hedge funds are subject to the caveat that the Jensen's alpha calculation ignores the differences in tax efficiency and, as we shall soon examine, liquidity provision between a hedge fund and its OE benchmark.

The table also reports style-based alphas, where the market index is replaced with a fund style-group specific benchmark. The style benchmark used for funds within a category is formed as a value-weighted portfolio of a sub-sample funds within that category. Benchmark returns are provided by Lipper/TASS. The style benchmarks explain much more of the variation in hedge fund returns as compared to the market index. The change in the average adjusted-R^2 ranges from -6% (EMN) to 35% (CA) across style groups. The style alphas are lower on average than the market-based alphas, and a greater fraction of funds deliver negative style alphas for all groups. Event Driven and Fund of Funds, in particular, look worse with a style benchmark. Overall, however, the average alpha is positive for most style groups.

The lagged market model is intended to reduce the potential bias in estimated alpha due to non-synchronous trading of the fund's underlying assets. The higher average adjusted-R^2 for most style groups indicates the presence of infrequent trading hedge fund assets. Also, the model delivers much lower average alphas for Convertible Arbitrage (0.28%–0.18%), Emerging Markets (0.48%–0.25%), and Fixed Income Arbitrage (0.24%–0.09%) style groups. ~~Asness et al. (2001) find a similar pattern across style.~~

The next measure examines market timing ability. The unconditional timing measure is the coefficient in the ~~Henriksson Merton~~ (1981) timing model (see Equation (2.6)). The average estimates are negative for seven of the eleven style groups, indicating that hedge funds market beta is actually lower during up-markets as compared to down-markets. However, the fractions of funds with negative timing coefficients are centered more closely around 50% across style groups, as compared to the mutual fund findings reported in the previous section. Taken together, the evidence is broadly consistent with much of the mutual fund literature, which often finds either neutral or poor timing performance.

The previous results ignore the illiquidity of a typical hedge fund share. Yet, many hedge funds impose restrictions on investor redemptions, such as lockups and redemption notice periods, thereby making hedge funds an illiquid investment. Investors may therefore expect higher returns on funds with share restrictions, commensurate with the illiquidity they bear. As discussed in Section 8.7.5, Aragon (2007) finds this to be true for hedge funds during 1994–2001. We follow Aragon (2007) and calculate liquidity-adjusted performance, equal to the raw performance less a liquidity premium given the fund's share restrictions. This involves a two-step procedure in which the performance estimates (e.g., Sharpe Ratio, Jensen's alpha) from the first step are used as dependent variables in the cross-section regression

$$a_p = \Pi_0 + \Pi_1 \mathrm{DLOCK}_p + \Pi_2 \mathrm{NOTICE}_p + \Pi_3 (\mathrm{NOTICE}_p)^2 + w_p, \tag{9.2}$$

where a_p is an estimate of fund performance, DLOCK is an indicator that equals one if the fund has a lockup provision and zero otherwise, and NOTICE is the number of days advance notice the fund requires an investor to redeem his shares. The coefficients could be interpreted as in Fama and MacBeth (1973) cross-sectional regressions, as premiums on share illiquidity factors. The presence of a quadratic term in Equation (9.2) is motivated by Amihud and Mendelson (1986). They argue that, if more illiquid assets are held by investors with longer investment horizons, then the relation between expected returns and illiquidity will be positive and concave.

Table 9.3 reports summary information for the liquidity-adjusted performance for each style group. On average, funds within the Managed Futures and Event Driven groups impose the lightest and heaviest liquidity restrictions, respectively. For example lockup usage is only 2% for MF, as compared to 40% for the ED funds. Meanwhile, the average ED fund requires about 48 days for share redemption as compared to only one week's notice for the MF category.

Consistent with Aragon (2007), we find a positive relation between share restrictions and performance. Liquidity-adjusted Sharpe Ratios are lower for every style group. Although the Sharpe Ratio drops from 0.05 to 0.03 for MF funds, the ED category experience a reduction of

Table 9.3 Summary of liquidity-adjusted hedge fund performance evidence.

Fund style	CA	DSB	EM	EMN	ED	FIA	FOF	GM	LS	MF	MS
Liquidity Variables											
Lockup?	0.25	0.25	0.17	0.22	0.40	0.23	0.18	0.10	0.33	0.02	0.29
Notice (days)	39.0	26.3	27.2	27.4	48.4	37.1	37.4	17.8	30.0	6.4	32.1
Sharpe Ratios											
Average	0.21	−0.14	0.12	0.06	0.15	0.28	0.11	0.01	0.05	0.03	0.17
% < Market	0.48	0.75	0.52	0.59	0.40	0.41	0.54	0.67	0.54	0.51	0.33
Jensen's Alphas											
Average (%)	−0.04	0.04	0.27	−0.04	0.10	−0.04	−0.10	−0.06	0.15	0.15	0.19
Percent < 0	0.54	0.54	0.32	0.47	0.37	0.36	0.51	0.42	0.37	0.30	0.33
Style Alphas											
Average (%)	−0.11	0.04	0.01	0.01	−0.26	0.00	−0.35	−0.49	−0.15	0.00	−0.02
Percent < 0	0.57	0.42	0.39	0.47	0.65	0.34	0.71	0.58	0.47	0.36	0.42
Lagged Market Model Alphas											
Average (%)	−0.21	0.13	−0.02	−0.12	−0.08	−0.27	−0.31	−0.29	−0.04	0.10	0.04
Percent < 0	0.64	0.58	0.38	0.58	0.48	0.47	0.66	0.51	0.45	0.32	0.46
Unconditional Timing											
Avg. coefficient	0.11	0.22	−0.67	0.05	−0.16	−0.03	−0.17	−0.09	−0.24	0.34	−0.15
Percent < 0	0.26	0.38	0.64	0.40	0.63	0.39	0.60	0.42	0.58	0.18	0.46

Note: The sample starts in January 1994 or later, depending on the fund, and ends in December 2005. The estimation excludes all return observations that precede the data the fund was added to the database. Fund styles are Convertible Arbitrage (CA), Dedicated Short Bias (DSB), Emerging Markets (EM), Equity Market Neutral (EMN), Event Driven (ED), Fixed Income Arbitrage (FIA), Fund of Funds (FOF), Global Macro (GM), Long/Short Equity (LS), Managed Futures (MF), and Multi-Strategy (MS). N is the number of funds with at least 24 monthly returns. Alphas are reported in percent per month. The lagged market model is the Jensen's alpha regression including three lags of the market excess return. The timing model is the Henriksson–Merton regression, where the coefficient on the positive part of the market excess return is summarized. The table summarizes liquidity-adjusted coefficients. The liquidity-adjusted coefficient equals $\text{adj}_p = a_p - [\Pi_1 \text{DLOCK}_p + \Pi_2 \text{NOTICE}_p + \Pi_3 (\text{NOTICE}_p)^2]$, where Π_1, Π_2, and Π_3 are the estimated coefficients of the following cross-sectional regression:

$$a_p = \Pi_0 + \Pi_1 \text{DLOCK}_p + \Pi_2 \text{NOTICE}_p + \Pi_3 (\text{NOTICE}_p)^2 + w_p,$$

where a_p is an estimate of fund p's performance, DLOCK is an indicator variable that equals one if the fund has a lockup, and NOTICE is the redemption notice period.

0.19 after controlling for share liquidity. The coefficients on the lockup (Π_1) and redemption notice (Π_2) variables are positive and significant. The proportion of liquidity-adjusted Sharpe Ratios falling below that of the market index is now centered at 50% across style groups. Thus, the liquidity adjustments appear to remove the evidence that hedge funds offer large Sharpe ratios.

In Table 9.3 the average liquidity-adjusted alphas are positive for only six of the hedge fund style groups, and they are lower for every category as compared to the unadjusted performance results. This can be explained by a 0.14% monthly lockup premium and a 0.25% average monthly premium per 30-day redemption notice period. The average liquidity-adjusted lagged market model alpha is *negative* for eight of the eleven style groups, and the average alpha across all funds is −0.12% per month. Overall, the evidence suggests that hedge funds do not deliver positive value-added to stock market investors, over and above the compensation for share restrictions.

Finally, we do not find a significant relation between hedge fund market timing ability and hedge fund share restrictions. This explains why the last rows in Table 9.3 are qualitatively similar to those in Table 9.2.

10

Conclusions

We have reviewed the models and methods for measuring portfolio performance and the evidence on the performance of professionally managed investment portfolios. Our review includes traditional measures, their properties and some of the important problems associated with the early measures. We also reviewed the more recent Conditional Performance Evaluation literature, weight-based performance measures and the stochastic discount factor approach, along with the evidence that these newer measures have produced. Our discussion has touched on equity style mutual funds, pension funds, asset allocation style funds, fixed income funds, and hedge funds. We draw several broad conclusions about the evidence that the literature has produced on fund performance, its relation to the efficiency of the markets, and also about future directions that we would like the literature to take.

We have defined investment performance at two broad levels. A fund or manager has investment ability if it generates returns that can be expected to exceed that of an otherwise equivalent benchmark, before costs and fees. But a fund may dissipate its ability through trading costs or capture the rents to its ability through management fees. A fund that outperforms the otherwise equivalent benchmark on an after-cost basis is said to add value for investors.

One underdeveloped area in the performance measurement literature is fixed income fund performance. There are many interesting methodological issues here, including which models and factors to use and the effects of interim trading and illiquidity on the performance measures. Finally, we think that future research needs to be more careful about the effects of costs and taxes in the evaluation of managed portfolio returns.

Acknowledgments

Ferson acknowledges financial support from the Collins Chair in Finance at Boston College and the Marshall School of Business at the University of Southern California. We are grateful to Stephen Brown, George Constantinides, Florian Munkel, and Min Kim for their helpful comments.

References

Ackermann, C., R. McEnally, and D. Ravenscraft (1999), 'The performance of hedge funds: Risk, return, and incentives'. *Journal of Finance* **54**, 833–874.

Admati, A. R., S. Bhattacharya, P. Pfleiderer, and S. A. Ross (1986), 'On timing and selectivity'. *Journal of Finance* **41**, 715–730.

Admati, A. R. and P. Pfleiderer (1997), 'Does it all add up? Benchmarks and the compensation of portfolio managers'. *Journal of Business* **70**, 323–350.

Agarwal, V. and N. Naik (2004), 'Risks and portfolio decisions involving hedge funds'. *Review of Financial Studies* **17**, 63–98.

Amihud, Y. and H. Mendelson (1986), 'Asset pricing and the bid-ask spread'. *Journal of Financial Economics* **17**, 223–249.

Aragon, G. (2005), ~~'Three essays on investments'. Unpublished Ph.D. Dissertation, Boston College.~~

Aragon, G. (2007), 'Share restrictions and asset pricing: Evidence from the hedge fund industry'. *Journal of Financial Economics* **83**, 33–58.

Aragon, G. and W. Ferson (2008), 'Evaluating hedge fund performance against dynamic nonlinear benchmarks'. Working Paper, Arizona State University.

Asness, C., R. Krail, and J. Liew (2001), 'Do hedge funds hedge?'. *Journal of Portfolio Management* **28**, 6–19.

Becker, C., W. Ferson, D. Myers, and M. Schill (1999), 'Conditional market timing with benchmark investors'. *Journal of Financial Economics* **52**, 119–148.

Berk, J. and R. C. Green (2004), 'Mutual fund flows and performance in rational markets'. *Journal of Political Economy* **112**, 1269–1295.

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

Black, F. and J. Treynor (1973), 'How to use security analysis to improve portfolio selection'. *Journal of Business* **46**, 66–86.

Brown, D. T. and W. J. Marshall (2001), 'Assessing fixed-income fund manager style and performance from historical returns'. *Journal of Fixed Income* **10**, 15–25.

Brown, K., V. Harlow, and L. Starks (1996), 'Of tournaments and temptations: An analysis of managerial incentives in the mutual fund industry'. *Journal of Finance* **51**, 85–110.

Brown, S., W. Goetzmann, and R. Ibbotson (1999), 'Offshore hedge funds: Survival and performance 1989–1995'. *Journal of Business* **72**, 91–118.

Brown, S. J. and W. N. Goetzmann (1995), 'Performance persistence'. *Journal of Finance* **50**, 679–98.

Brown, S. J., W. N. Goetzmann, R. Ibbotson, and S. Ross (1992), 'Survivorship bias in performance studies'. *Review of Financial Studies* **5**, 553–580.

Busse, J. (1999), 'Volatility timing in mutual funds: Evidence from daily returns'. *Review of Financial Studies* **12**, 1009–1041.

Busse, J. (2001), 'Another look at mutual fund tournaments'. *Journal of Financial and Quantiative Analysis* **36**, 53–73.

Carhart, M., J. Carpenter, A. Lynch, and D. Musto (2002), 'Mutual fund survivorship'. *Review of Financial Studies* **15**, 1439–1463.

Carhart, M. M. (1997), 'On persistence in mutual fund performance'. *Journal of Finance* **52**, 57–82.

Carlson, R. S. (1970), 'Aggregate performance of mutual funds, 1948–1967'. ~~*The*~~ *Journal of Financial and Quantitative Analysis* **5**, 1–32.

Chang, E. C. and W. G. Lewellen (1984), ‘Market timing and mutual fund investment performance’. *Journal of Business* **57**, 57–72.

Chen, N.-F., T. Copeland, and D. Mayers (1987), ‘A comparison of single and multifactor perforamance methodologies’. *Journal of Financial and Quantitative Analysis* **22**, 401–417.

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

Chen, Y., W. Ferson, and H. Peters (2006), ‘Measuring the timing ability of fixed income funds’. Unpublished Working Paper, University of Southern California.

Chen, Z. and P. J. Knez (1996), ‘Portfolio performance Measurement: Theory and applications’. *Review of Financial Studies* **9**, 511–556.

Chevalier, J. and L. Ellison (1997), ‘Risk taking by mutual funds as a response to incentives’. *Journal of Political Economy* **105**, 1167–1200.

Chiu, S. and V. Roley (1992), ‘Risk incentives in mutual fund management: Theory and empirical results’. Unpublished Working Paper, University of Washington, Seattle.

Chordia, T. (1996), ‘The structure of mutual fund charges’. *Journal of Financial Economics* **41**, 3–39.

Christopherson, J. A., W. Ferson, and D. A. Glassman (1998a), ‘Conditional measures of performance and persistence for pension funds’. In: *Research in Finance*, Vol. 16. Stamford, CT: JAI Press, pp. 1–46. ISBN: 0-7623-0328-X.

Christopherson, J. A., W. Ferson, and D. A. Glassman (1998b), ‘Conditioning manager alpha on economic information: Another look at the persistence of performance’. *Review of Financial Studies* **11**, 111–142.

Christopherson, J. A. and A. L. Turner (1991), ‘Volatility and predictability of manager alpha: Learning the lessons of history’. *Journal of Portfolio Management* **Fall**, 5–12.

Coggin, D., F. Fabozzi, and S. Rahman (1993), ‘The investment performance of U.S. equity pension fund managers’. *Journal of Finance* **48**, 1039–1056.

Comer, G. (2006), ‘Evaluating bond fund sector timing skill’. Unpublished Working Paper, Georgetown University.

Comer, G., V. Boney, and L. Kelly (2005), 'Timing the investment grade securities market: Evidence from high quality bond funds'. Unpublished Working Paper, Georgetown University.

Connor, G. and R. A. Korajczyk (1986), 'Performance measurement with the arbitrage pricing theory: A new framework for analysis'. *Journal of Financial Economics* **15**, 373–394.

Connor, G. and R. A. Korajczyk (1988), 'Risk and return in an equilibrium APT: Applications of a new test methodology'. *Journal of Financial Economics* **21**, 255–289.

Copeland, T. and D. Mayers (1982), 'The value line enigma (1965–1978): A case study of performance evaluation issues'. *Journal of Financial Economics* **10**, 289–321.

Cornell, B. (1979), 'Asymmetric information and portfolio performance measurement'. *Journal of Financial Economics* **7**, 381–390.

Cox, J. C., J. E. Ingersoll Jr., and S. A. Ross (1985), 'A theory of the term structure of interest rates'. *Econometrica* **53**, 385–346.

Dahlquist, M. and P. Soderlind (1999), 'Evaluating portfolio performance with stochastic discount factors'. *Journal of Business* **72**, 347–384.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers (1997), 'Measuring mutual fund performance with characteristic-based benchmarks'. *Journal of Finance* **52**, 1035–1058.

Del Guercio, D. and P. A. Tkac (2002), 'The determinants of the flow of funds of managed portfolios: Mutual funds vs. Pension funds'. *Journal of Financial and Quantitative Analysis* **37**, 523–557.

Dimson, E. (1979), 'Risk measurement when shares are subject to infrequent trading'. *Journal of Financial Economics* **7**, 197–226.

Dybvig, P. H. and J. E. Ingersoll (1982), 'Mean variance theory in complete markets'. *Journal of Business* **55**, 233–251.

Dybvig, P. H. and S. A. Ross (1985), 'Performance measurement using differential information and a security market line'. *Journal of Finance* **40**, 383–399.

Edelen, R. M. (1999), 'Investor flows and the assessed performance of open-end mutual funds'. *Journal of Financial Economics* **53**, 439–466.

Elton, E. J., M. Gruber, and C. Blake (2001), 'A first look at the accuracy of the CRSP mutual fund database and a comparison of the CRSP and Morningstar databases'. *Journal of Finance* **56**, 2415–2430.

Elton, E. J., M. J. Gruber, and C. R. Blake (1993), 'The performance of bond mutual funds'. *Journal of Business* **66**, 371–403.

Elton, E. J., M. J. Gruber, and C. R. Blake (1995), 'Fundamental economic variables, expected returns and bond fund performance'. *Journal of Finance* **50**, 1229–1256.

Elton, E. J., M. J. Gruber, and C. R. Blake (1996), 'Survivorship bias and mutual fund performance'. *Review of Financial Studies* **9**, 1097–1120.

Elton, E. J., M. J. Gruber, and C. R. Blake (2003), 'Incentive fees and mutual funds'. *Journal of Finance* **58**, 779–804.

Evans, R. B. (2003), 'Does alpha really matter? Evidence from fund incubation, termination and manager change'. Working Paper, University of Pennsylvania.

Fama, E. (1972), 'Components of investment performance'. *Journal of Finance* **27**, 551–567.

Fama, E. F. (1970), 'Efficient capital markets: A review of theory and empirical work'. *Journal of Finance* **25**, 383–417.

Fama, E. F. and K. R. French (1996), 'Multifactor explanations of asset pricing anomalies'. *Journal of Finance* **51**, 55–87.

Fama, E. F. and J. D. MacBeth (1973), 'Risk, return, and equilibrium: Some empirical tests'. *Journal of Political Economy* **81**, 607–636.

Farnsworth, H. K. (1997), 'Evaluating stochastic discount factors from term structure models'. Unpublished Ph.D. Dissertation, University of Washington.

Farnsworth, H. K., W. Ferson, D. Jackson, and S. Todd (2002), 'Performance evaluation with stochastic discount factors'. *Journal of Business* **75**, 473–504.

Ferson, W., T. Henry, and D. Kisgen (2006a), 'Evaluating government bond fund performance with stochastic discount factors'. *Review of Financial Studies* **19**, 423–455.

Ferson, W., T. Henry, and D. Kisgen (2006b), 'Fixed income fund performance across economic states'. In: *Research in Finance*, Vol. 23. Oxford, UK: JAI Press, pp. 1–62. ISBN-13: 978-0-7623-1346-7.

Ferson, W. and K. Khang (2002), 'Conditional performance measurement using portfolio weights: Evidence for pension funds'. *Journal of Financial Economics* **65**, 249–282.

Ferson, W. and M. Qian (2004), 'Conditional performance evaluation revisited'. AIMR manuscript.

Ferson, W. and A. F. Siegel (2001), 'The efficient use of conditioning information in portfolios'. *Journal of Finance* **56**, 967–982.

Ferson, W. and A. F. Siegel (2006), 'Testing portfolio efficiency with conditioning information'. Working Paper, University of Southern California.

Ferson, W. and V. A. Warther (1996), 'Evaluating fund performance in a dynamic market'. *Financial Analysts Journal* **52**, 20–28.

Ferson, W. E. (1995), 'Theory and empirical testing of asset pricing models, Chapter 5 in Finance'. In: M. Jarrow and Ziemba (eds.): *Handbooks in Operations Research and Management Science*. Elsevier, pp. 145–200.

Ferson, W. E. (2003), 'Tests of multifactor pricing models, volatility bounds and portfolio performance'. In: M. H. George M. Constantinides and R. M. Stulz (eds.): *Handbook of the Economics of Finance*. North Holland: Elsevier Science Publishers (forthcoming).

Ferson, W. E., S. Sarkissian, and T. Simin (2007), 'Asset pricing models with conditional betas and alphas: The effects of data mining and spurious regression'. *Journal of Financial and Quantitative Analysis*. forthcoming.

Fung, W. and D. Hsieh (1997), 'Empirical characteristics of dynamic trading strategies: The case of hedge funds'. *Review of Financial Studies* **10**, 275–302.

Fung, W. and D. Hsieh (2000), 'Performance characteristics of hedge funds and commodity funds: Natural vs. spurious biases'. *Journal of Financial and Quantitative Analysis* **35**, 291–307.

Fung, W. and D. Hsieh (2001), 'The risk in hedge fund strategies: Theory and evidence from trend followers'. *Review of Financial Studies* **14**, 313–341.

Getmansky, M., A. Lo, and I. Makarov (2004), 'An econometric model of serial correlation and illiquidity in hedge fund returns'. *Journal of Financial Economics* **74**, 529–610.

Glosten, L. and R. Jagannathan (1994), 'A contingent claims approach to performance evaluation'. *Journal of Empirical Finance* **1**, 133–166.

Goetzmann, W., J. Ingersoll, and Z. Ivkovic (2000), 'Monthly measurement of daily timers'. *Journal of Financial and Quantitative Analysis* **35**, 257–290.

Goetzmann, W., J. Ingersoll, M. Spiegel, and I. Welch (2005), 'Sharpening sharpe ratios'. NBER Working Paper No. 9116.

Goetzmann, W. N. and R. G. Ibbotson (1994), 'Do winners repeat?'. *Journal of Portfolio Management* **20**, 9–18.

Goriaev, M., T. Nijman, and B. Werker (2005), 'Yet another look at mutual fund tournaments'. *Journal of Empirical Finance* **12**, 127–137.

Grant, D. (1977), 'Portfolio performance and the 'cost' of timing decisions'. *Journal of Finance* **32**, 837–846.

Grinblatt, M. and S. Titman (1989a), 'Mutual fund performance: An analysis of quarterly portfolio holdings'. *Journal of Business* **62**, 393–416.

Grinblatt, M. and S. Titman (1989b), 'Portfolio Performance evaluation: Old issues and new insights'. *Review of Financial Studies* **2**, 393–422.

Grinblatt, M. and S. Titman (1992), 'The persistence of mutual fund performance'. *Journal of Finance* **47**, 1977–1984.

Grinblatt, M. and S. Titman (1993), 'Performance measurement without benchmarks: An examination of mutual fund returns'. *Journal of Business* **60**, 97–112.

Grinblatt, M., S. Titman, and R. Wermers (1995), 'Momentum strategies, portfolio performance and herding: A study of mutual fund behavior'. *American Economic Review* **85**, 1088–1105.

Grossman, S. J. and J. E. Stiglitz (1980), 'On the impossibility of informationally efficient markets'. *American Economic Review* **70**, 393–408.

Gruber, M. (1996), 'Another puzzle: The growth in actively managed mutual funds'. *Journal of Finance* **51**, 783–810.

Hansen, L. P. (1982), 'Large sample properties of generalized method of moments estimators'. *Econometrica* **50**, 1029–1054.

Hansen, L. P. and S. Richard (1987), 'The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models'. *Econometrica* **55**, 587–613.

Heinkel, R. and N. M. Stoughton (1995), 'A new method for portfolio performance measurement'. Unpublished Working Paper, University of British Columbia.

Hendricks, D., J. Patel, and R. Zeckhauser (1993), 'Hot hands in mutual funds: Short run persistence of performance, 1974–1988'. *Journal of Finance* **48**, 93–130.

Hendricks, D., J. Patel, and R. Zeckhauser (1997), 'The J-shape of performance persistence given survivorship bias'. *Review of Economics and Statistics* **79**, 161–166.

Henriksson, R. (1984), 'Market timing and mutual fund performance: An empirical Investigation'. *Journal of Business* **57**, 73–96.

Ippolito, R. A. (1989), 'Efficiency with costly information: A study of mutual fund performance, 1965–1984'. *The Quarterly Journal of Economics* **104**, 1–23.

Ippolito, R. A. (1992), 'Consumer reaction to measures of poor quality: Evidence from the mutual fund industry'. *Journal of Law and Economics* **35**, 45–70.

Jagannathan, R. and R. Korajczyk (1986), 'Assessing the market timing performance of managed portfolios'. *Journal of Business* **59**, 217–236.

Jensen, M. C. (1968), 'The performance of mutual funds in the period 1945-1964'. *Journal of Finance* **23**, 389–416.

Jensen, M. C. (1969), 'Risk, the pricing of capital assets, and the evaluation of investment portfolios'. *Journal of Business* **42**, 167–247.

Jensen, M. C. (1972), 'Optimal utilization of market forecasts and the evaluation of investment performance'. In: G. P. Szego and K. Shell (eds.): *Mathematical Methods in Finance.* North-Holland Publishing Company.

Kon, S. J. (1983), 'The market timing performance of mutual fund managers'. *Journal of Business* **56**, 323–347.

Koski, J. and J. Pontiff (1999), 'How are derivatives used? Evidence from the mutual fund industry'. *Journal of Finance* **54**, 791–816.

Kryzanowski, L., S. Lalancette, and M. To (1997), 'Performance attribution using an APT with prespecified factors'. *Journal of Financial and Quantitative Analysis* **32**, 205–224.

Lakonishok, J., A. Shleifer, and R. Vishny (1992), 'The structure and performance of the money management industry'. *Brookings Papers on Economic Activity: Microeconomics,* pp. 339–391.

Lehmann, B. N. and D. M. Modest (1987), 'Mutual fund performance evaluation: A comparison of benchmarks and benchmark comparisons'. *Journal of Finance* **42**, 233–266.

Leland, H. (1999), 'Performance beyond mean-variance: Performance measurement in a nonsymmetric world'. *Financial Analysts Journal* **55**, 27–36.

Liang, B. (1999), 'On the performance of hedge funds'. *Financial Analysts Journal* **55**, 72–85.

Liang, B. (2000), 'Hedge funds: The living and the dead'. *Journal of Financial and Quantitative Analysis* **35**, 309–326.

Long, J. B. (1974), 'Stock prices, inflation, and the term structure of interest rates'. *Journal of Financial Economics* **1**, 131–170.

Malkiel, B. G. (1995), 'Returns from investing in equity mutual funds 1971 to 1991'. *Journal of Finance* **50**, 549–572.

Mayers, D. and E. M. Rice (1979), 'Measuring portfolio performance and the empirical content of asset pricing models'. *Journal of Financial Economics* **7**, 3–28.

~~Ferson, W. and R. Schadt (1996), 'Measuring fund strategy and performance in changing economic conditions'. *Journal of Finance* **51**, 425–462.~~

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

Merton, R. C. (1981), 'On market timing and investment performance I: An equilibrium theory of value for market forecasts'. *Journal of Business* **54**, 363–406.

Merton, R. C. and R. D. Henriksson (1981), 'On market timing and investment performance II: Statistical procedures for evaluating forecasting skills'. *Journal of Business* **54**, 513–534.

Mitchell, M. and T. Pulvino (2001), 'Characteristics of risk and return in risk arbitrage'. *Journal of Finance* **56**, 2135–2175.

Park, J. M. (1995), 'Managed futures as an investment class asset'. Unpublished Ph.D. Dissertation, Columbia University.

Posthuma, N. and P. van der Sluis (2003), 'A reality check on hedge fund returns'. Unpublished Working Paper, Free University Amsterdam.

Roll, R. (1978), 'Ambiguity when performance is measured by the security market line'. *Journal of Finance* **33**, 1051–1069.

Ross, S. A. (1976), 'The arbitrage pricing theory of capital asset pricing'. *Journal of Economic Theory* **13**, 341–360.

Scholes, M. and J. Williams (1977), 'Estimating betas from nonsynchronous data'. *Journal of Financial Economics* **5**, 309–327.

Schultz, E. (1996), 'Vanguard bucks trend by cutting fund fees'. *The Wall Street Journal,* pp. c1–c25.

Sharpe, W. F. (1964), 'Capital asset prices: A theory of market equilibrium under conditions of risk'. *Journal of Finance* **19**, 425–442.

Sharpe, W. F. (1966), 'Mutual fund performance'. *Journal of Business* **1**(part II), 119–138.

Sharpe, W. F. (1992), 'Asset allocation: Management style and performance measurement'. *Journal of Portfolio Management* **18**, 7–19.

Shukla, R. and C. Trzcinka (1994), 'Persistent performance in the mutual fund market: Tests with funds and investment advisors'. *Review of Quantitative Finance and Accounting* **4**, 115–135.

Sirri, E. R. and P. Tufano (1998), 'Costly search and mutual fund flows'. *Journal of Finance* **53**, 1589–1622.

Stanton, R. (1997), 'A nonparametric model of term structure dynamics and the market price of interest rate risk'. *Journal of Finance* **52**, 1973–2002.

Starks, L. (1987), 'Performance incentive fees: An agency theoretic approach'. *Journal of Financial and Quantitative Analysis* **22**, 17–32.

Treynor, J. (1965), 'How to rate management of investment funds'. *Harvard Business Review* ~~(January–February), pp.~~ 63–75.

Treynor, J. and K. Mazuy (1966), 'Can mutual funds outguess the market?'. *Harvard Business Review* **44**, 131–136.

Vasicek, O. A. (1977), 'An equilibrium characterization of the term structure'. *Journal of Financial Economics* **5**, 177–188.

Verrecchia, R. E. (1980), 'Consensus beliefs, information acquisition, and market information efficiency'. *American Economic Review* **70**, 874–884.

Zheng, L. (1999), 'Is money smart? A study of mutual fund investors' fund selection ability'. *Journal of Finance* **54**, 901–933.